

Enterprise
Social Media Management
baseret på
Data Mining af Tweets

Ulrik Uhre Brink - s081830
&
Kristian Michael Clarkson - s072660

DTU



Kongens Lyngby 2014

Danmarks Tekniske Universitet
DTU Compute - Institut for Matematik og Computer Science
Bygning 324, DK-2800 Kongens Lyngby, Denmark
Telefon +45 45253031, Fax +45 45881399
compute@compute.dtu.dk
www.compute.dtu.dk

Forord

Denne afhandling er udarbejdet ved DTU Compute - Institut for Matematik og Computer Science på Danmarks Tekniske Universitet i opfyldelse af kravene for erhvervelse af en M.Sc. i Digitale Medieteknologier.

Afhandlingen beskæftiger sig med data mining af tweets fra Twitter, herunder filtrering og sortering, samt design og udarbejdelsen af en web-applikation.

Afhandlingen består af et leveret produkt, i form af en web-applikation, samt en rapport til dokumentation af denne.

Vi vil gerne takke vores vejleder på DTU, Michael Kai Petersen, for tid og vejledning under projektføreløbet.

Lyngby, 28-Juni-2014

Two handwritten signatures in black ink. The signature on the left is 'KMc' and the signature on the right is 'UB'.

Kristian Michael Clarkson & Ulrik Uhre Brink

Resumé

Social data er blevet en vigtig del ved lancering af et nyt produkt og anvendes især til at afholde kommunikationen med dets modtagere. De sociale medier har skabt endnu en kommunikationskanal, og igennem denne er det muligt at få indblik i brugerne og deres interaktion imellem. Blot på det sociale medie Twitter, findes der flere hundrede millioner af tweets, og antallet er stærkt stigende. Dette skaber uoverskuelighed, og det kan være svært for virksomheder at finde og vælge relevante informationer i mængden af social data.

Projektet vil blive kørt som et lean-startup og i den forbindelse vil der blive udarbejdet et produkt samt en forretningsmodel for dette. Produktet vil have form som en web-applikation, hvis formål er at præsentere analyseret social data, som er relevante for virksomhederne. På denne måde skabes overblik på et ellers uoverskueligt marked. Gennem en iterativ proces undervejs i projektet, vil forretningsmodellen blive evalueret på baggrund af erfaringer og ny viden undervejs.

Det vil i projektet forsøges at data mine Twitter for social data, for at kunne afgrænse og finde relevant information. På den måde vil det være muligt at omdanne kvantitativ data til kvalitativ data, som virksomheder kan gøre brug af.

At indhente relevant information fra social data er en længere proces, som kræver forskellige typer af delementer, der alle er afhængige af hinanden. Gennem forskellige filtreringsmetoder vil det blive muligt at frasortere hvid støj, herunder ikke-relevante tweets såsom reklamer etc. Ved frasortering af hvid støj, vil data være gjort klar til videre analyse, således at relevante data kan indhentes. Nogle tweets er mere relevante for virksomheden end andre, og gennem forskellige sorteringsmetoder, herunder rangering af brugere, stemningsanalyse etc., vil de mest relevante tweets blive fremhævet. Disse analyser vil også kunne bestemme den generelle holdning til virksomhedens brand.

Igennem to cases, *Eurovision* og *Voteman*, vil yderligere analyse forsøge at illustrere emnerne ud fra forskellige aspekter. Ved *Eurovision* vil det forsøges, at forudsige vinderen af årets melodi grandprix, og hvilke andre emner der ud-

sprang herfra. Ved *Voteman* vil det forsøges at finde dets oprindelse, og illustrere hvordan det gennem forløbet har spredt sig til hele verden.

Det udarbejdede produkt vil forsøges promoveret gennem en landing-page og en opsat twitter-profil. Baseret på opsatte *call-to-actions* vil det være muligt at indhente informationer omkring produktets efterspørgsel på markedet og dermed afgøre, hvorvidt man bør gå videre med produktet eller ej.

Indhold

Forord	i
Resumé	iii
1 Indledning	1
2 Sociale Medier	3
2.1 Definition af sociale medier	3
2.2 Effekten af de sociale medier	4
2.3 Kendetegn ved de største sociale netværk	5
2.4 De sociale medier i tal	7
2.5 Evaluering af sociale medier	8
3 Markedsanalyse	9
3.1 Det digitale marked	9
3.1.1 Forretningsmodeller for digitale produkter	10
3.2 Konkurrenter	11
3.3 Evaluering af markedsanalyse	12
4 Forretningsmodel	15
4.1 Hvorfor lean startup?	15
4.1.1 Build - Measure - Learn	15
4.2 Lean forretningsmodel	16
4.2.1 Lean Canvas Model	17
4.3 Evaluering af forretningsmodel	25
5 1. iteration	29
5.1 Twitter	29
5.2 Opstilling af hypoteser	30
5.3 Evaluering af 1. iteration	31
6 2. iteration	33
6.1 Hypoteser	33
6.2 Indsamling af data	33
6.3 Forbindelse til Twitter API	34
6.4 Analyse af indhentet data	35
6.5 Evaluering af 2. iteration	37

6.6	Validering af forretningsmodel	37
7	3. iteration	39
7.1	Hypoteser	39
7.2	Brand-filtrering	39
7.3	Frasortering af ikke relevant data	40
7.4	Evaluering af 3. iteration	45
7.5	Validering af forretningsmodel	45
8	4. iteration	47
8.1	Hypoteser	47
8.2	Frekvensfordeling	48
8.3	Stemmingsanalyse	50
8.3.1	AFINN	51
8.3.2	Naive Bayes	51
8.4	Rangering og Sortering	53
8.4.1	En relevant bruger	53
8.4.2	En vigtig tweet	55
8.5	Evaluering af 4. iteration	57
8.6	Validering af forretningsmodel	58
9	5. iteration	59
9.1	Hypoteser	59
9.2	To typer af MVP	59
9.3	Landing-page	60
9.3.1	Landing-page cases	61
9.3.2	Gennemgang af landing-page	61
9.3.3	Indhentning af data	64
9.4	Web-applikation	66
9.4.1	Gennemgang af web-applikation	67
9.4.2	Evaluering og validering af web-applikation	71
9.5	Evaluering af 5. iteration	73
9.6	Validering af forretningsmodel	75
10	6. iteration	77
10.1	Social data analyse	77
10.1.1	Voteman	78
10.1.2	Eurovision	84
10.2	Evaluering af 6. iteration	91
10.3	Validering af forretningsmodel	93
11	Konklusion	95
	Referencer	97

Bilag	105
A Google Cloud omkostninger	105
B TV2 Breaking News Tabel	106
C #twitterhjerne eksempler	107
D Stopordsliste	108
E Iteration fire - Wordclouds	111
F Landing page cases	114
G Landing page - Tagster.dk	115
H Landing page - Resultater	118
I Web-applikation - Tagster.dk	122
J Web-applikation - betatest	124
K Case - Voteman	127
L Case - Voteman - Valgfi	130
M Case - Voteman - Instagram/Twitter	131

Indledning

Særligt de sociale medier har ændret den måde man kommunikerer på[22]. Fysiske afstande er blevet brudt, og fremfor kun at kommunikere med venner og bekendte gør sociale medier det muligt at kommunikere med personer verden over. De sociale medier har også gjort det nemmere for personer at komme ud med meninger og holdninger til en bred skare af brugere, omkring givne emner.

Især under store events anvendes sociale medier, hvor personer her kan diskutere og debattere live begivenheder[66]. Dette kan betegnes som *second screen*, hvor tv bliver *set* via de sociale medier. Personer samles eksempelvis omkring et *hashtag* og diskuterer løbende begivenheder, mens de udfolder sig. Disse events skaber utroligt store mængder data, og alene under Eurovision-finalen på Twitter, blev der skrevet over 5,3 mio. tweets[76].

Brugerne på de sociale medier bidrager med at skabe store mængder data[50], og gør det lukrativt for virksomheder at lytte med i samtaler. Her kan der indhentes både meninger og holdninger omkring emner, men i særdeleshed også hvad der blandt brugerne trender.

Data er ustruktureret[81] og uden de rette værktøjer kan virksomheder *drukne* i mængden af brugergeneret data. Denne udfordring vil blive forsøgt løst med en web-applikation, hvis hovedfokus er at indhente, analysere og skabe overblik på et ellers uoverskueligt marked. Gennem analyser vil det indhentede data blive filtreret, sorteret og rangeret, således at det mest relevante vil blive præsenteret.

Sociale Medier

De sociale medier er kommet for at blive, og virksomheder bør gøre brug af de store mængder data, som er frit tilgængelige. Dette kapitel beskriver en definition af sociale medier, hvordan de sociale medier har ændret den måde vi arbejder på, samt en analyse af de største sociale netværk og deres kendetegn.

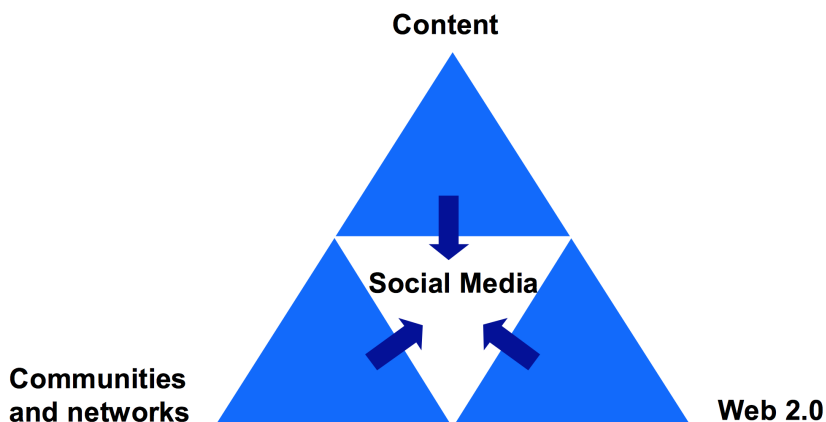
2.1 Definition af sociale medier

Det kan være svært at give en præcis definition af hvad sociale medier er, og ifølge T. Ahlqvist[1] er et socialt medie bygget op omkring tre nøgle områder, også beskrevet som den sociale medie trekant - *indhold, fællesskab og netværk* samt udnyttelsen af *Web 2.0 teknologi*. Den sociale medie trekant er illustreret i figur 2.1.

Den første del af den sociale medie trekant er *indholdet*. Indholdet er genereret af brugerne og kan eksistere i forskellige former alt afhængigt, hvilken kontekst det er skabt i. Det kan eksempelvis eksistere som rå tekst i en simpel status opdatering på Facebook[26], som en tweet på Twitter[73], som et billede eller video delt på Instagram[40].

De forskellige typer af brugergenereret indhold består af, alt afhængigt af indholdet, værdifulde informationer, som man vil kunne indsamle og analysere til eksempelvis at finde kundernes holdning til et givent produkt.

Den anden del af den sociale medie trekant er *fællesskab og netværk*, hvilket understreges af begrebet social. Uden den sociale del i fællesskabet ville det ikke være interessant for brugerne at skabe og dele indhold i disse netværk. Netværkene har ingen grænser, idet de tillader brugerne i netværket at dele



Figur 2.1: Den social medie trekant, kilde: [1, s. 14]

indhold med andre brugere, uanset om de kender de andre brugere eller ej. Det er den sociale faktor, der driver disse netværk, da brugerne har en vis trang til at dele særlige øjeblikke med andre på disse sociale netværk.

Den sidste del i den sociale medie trekant er *Web 2.0 teknologier*, der gør det muligt at oprette disse sociale netværk herunder Facebook[26], hvor brugerne kan skabe og dele deres unikke øjeblikke med andre brugere. De sociale netværk bygget op omkring Web 2.0, har ændret den måde man kommunikerer på idag både brugerne imellem, men også mellem en kunde og virksomhed. Teknologierne gør det muligt for brugerne, ikke alene at kunne læse og hente oplysninger om bestemte emner, men også **bidrage**, **kommentere** og **dele** disse med andre - især disse tre aspekter er vigtige på de sociale medier.

2.2 Effekten af de sociale medier

De sociale medier har i den grad sat sine spor på mange områder, men især på to områder menes det, at de sociale medier har haft den største indflydelse. Det ene er interaktionen brugerne imellem, og den anden er den måde nyhederne anvender de sociale medier på, som Benjamin Rud Elberth - Digital (politisk) kommunikator, også forklarer.

„Historier bliver plantet på Twitter. Historier bliver fundet på Twitter. Historier bliver forhandlet på Twitter, og historier bliver spredt på Twitter. Og endnu større – er dit brand eller CEO ikke tilgængelig på Twitter, står du sidst i køen

til at komme på citat på nyhedsfladerne“ [24], Elberth (2013).

Følgende er en gennemgang af de to områder, samt eksempler på hvorfor netop disse har fået indflydelse fra de sociale medier.

Nyhederne har ændret sig efter de sociale medier er kommet til. Tidligere var kilder til nyheder noget, som journalister skulle opsøge for at få den gode historie. Idag behøver journalister ikke forlade kontoret for at få den gode historie. Nyheden om hvad folk på Twitter stemmer vedr. eksempelvis patentdomstolen[37], blev via Twitter bragt som nyhed på nyhedssiden Version2[80]. Dette skete formentlig på baggrund af, at DTU¹ valgte at dele den studerendes tweet[38], idet den studerende ikke selv er blevet kontaktet af journalisten bag nyheden.

Mange nyhedsbureauer har valgt at oprette en særskilt "breaking news"profil, som skal bringe de vigtige nyheder ud til der hvor folket er. Et godt eksempel er TV2Breaking[58] på Twitter, hvor der bliver bragt live nyheder ud til deres over 43.000² følgere. Deres ti seneste nyheder er blevet delt af 60 følgere, som er vist i Bilag B.

Interaktionen blandt personer har også ændret sig siden de sociale medier kom til. De nye sociale netværk gør det muligt at holde kontakten, men også finde nye bekendtskaber, hvor man deler samme interesse. Det populære hashtag, #Twitterhjerne, er et glimrende eksempel på hvordan man, fremfor at søge Google tomt for søgeresultater, spørger og får hjælp fra andre personer. Eksempler på disse er illustreret figur 2 i Bilag C, og illustrerer hvordan det sociale medie bidrager til besvarelse af folks spørgsmål - uden at man faktisk kender de andre brugere.

2.3 Kendetegn ved de største sociale netværk

De sociale medier er efterhånden overalt, og det er svært at finde en hjemmeside, hvor man ikke har integreret enten Facebook's "synes godt om" eller Twitter's "tweet dette".

Selvom de sociale netværk er at finde overalt, kan det til tider være svært at få et overblik over, hvad der kendetegner dem hver især, hvor vigtige de er, og hvordan de er forskellige fra hinanden. Følgende er en gennemgang og analyse af de største sociale netværk, hvor man kan indhente relevante data. Analysen

¹Danmarks Teknisk Universitet

²43.041 følgere d. 10-06-2014

indeholder desuden en kort beskrivelse af deres særlige kendetegn, et overblik over det data som er til rådighed, samt eventuelle begrænsninger.

Twitter[73] kendetegnes bedst som mindre opdateringer i realtid og især anvendelsen af hashtags. En opdatering på Twitter betegnes som en tweet, og kan indeholde både tekst og billede. Det er dog begrænset til 140 tegn. Begrænsningen gør, at en tweet antages for at være mere velovervejet pga plads-restriktionen, men denne begrænsning er også med til at gøre det sværere at indhente relevant information ud fra de få ord en tweet indeholder. Anvendelsen af hashtags gør det muligt at finde tweets, hvor brugere aktivt deltager i debatten omkring et givent emne.

Facebook[26] kendetegnes bedst ud fra opdateringer omkring livsbegivenheder, og især sider/kunstnere etc. som man kan "synes godt om". Opdateringer på Facebook kan indeholde både tekst, billeder og video, men også mere kontekstuelle opdateringer. Facebook Open Graph gør det muligt at foretage opdateringer i form af "triples". En triple kan eksempelvis indeholde information om, hvem brugeren er sammen med, og hvor de er "checket ind". Disse informationer bidrager positivt, når der skal indhentes relevant information ud fra eksempelvis en opdatering fra en bruger. Facebook har ligesom Twitter gjort det muligt, at anvende hashtags i opdateringer, som gør det muligt at finde opdateringer fra forskellige brugere om det samme emne. Facebook og dets brugere har den begrænsning, at i langt de fleste tilfælde, bliver opdateringer kun sendt ud til brugere, som man er i netværk med. Dette besværliggør processen for data mining af hashtags, idet kun et begrænset antal opdateringer vil blive vist.

LinkedIn[46] kendetegnes bedst som et forretningsorienteret netværk, hvor fokus er på det forretningsmæssige aspekt herunder personers erfaringer, kompetencer og baggrund. Opdateringer omhandler i langt de fleste tilfælde, modsat Twitter og Facebook, emner der er relateret til ens baggrund eller kompetence-sæt. LinkedIn giver mulighed for at indhente information omkring en brugers forretningsmæssige baggrund, hvem brugeren er i netværk med, hvad de har tilfælles, samt hvilke virksomheder brugeren følger.

Google+[35] er bedre kendt som Google's eget sociale netværk. Her anvendes cirkler til inddeling af netværk, og +1 der er en pendant til Facebook's "synes godt om". Netværkets stærke side er dets indhold, idet der modsat Twitter og Facebook er langt færre "reklamer" og "støj", men dette understøttes også af, at der er langt færre brands aktive på Google+[67], og det antages derfor at det er pga. den niche gruppe af personer, som er af finde på netværket.

Instagram[40] kendetegnes bedst som det netværk, hvor flest billeder og videoer bliver delt i realtid. Billederne understøttes af en kommentar, hvilket bidrager positivt som en feature til billedet. Der anvendes hashtags i stor stil, som gør

det til et stærkt medie hvad angår data-mining, hvilket gør det nemt at finde billeder, som er relevante indenfor et bestemt emne. Dog ses det ofte, at brugere anvender flere hashtags end hvad godt er, og derfor er det svært at bestemme hvilket hashtag, som beskriver billedet bedst.

2.4 De sociale medier i tal

De sociale medier har stor indflydelse på personers gøren og laden, men hvor meget data der virkelig er til rådighed, kan kun de seneste tal fra statistikker fortælle. Med udgangspunkt i de mest anvendte og velkendte sociale netværk fundet i afsnit 2.3, er følgende en oversigt over hvor meget data der faktisk bliver genereret.

Følgende data er indhentet fra Leverage New Age Media[50], som har indsamlet data fra de største sociale netværk, og på baggrund af disse udarbejdet en oversigt.

Twitter har ud af deres i alt 600 millioner brugere globalt set over 255 millioner aktive brugere på månedlig basis. Disse brugere bidrager hvert sekund med over 5.700 tweets, hvilket på månedlig basis giver mere end 14 milliarder³ tweets. Twitter er mest udbredt i USA, hvor 18%⁴ af den amerikanske befolkning har en aktiv twitter-profil. I Danmark er Twitter ikke nær så udbredt, og ifølge Overskrift.dk[9] er der ca. en kvart million aktive twitter-profiler, antallet er dog stærkt stigende.

Facebook er med deres over 1 milliard brugere, globalt set, det største sociale netværk. Disse brugere bidrager hvert 20. minut med over 1 million opdateringer, hvilket på månedlig basis giver ca. 2 milliarder⁵ opdateringer.

LinkedIn har globalt set over 300 millioner brugere, hvoraf 200 millioner af dem er brugere udenfor USA. Fokus på LinkedIn er profiler samt jobopslag, og ifølge data fra LinkedIn[45], er der ca. 15 millioner profil visninger, over en million visninger af jobopslag, fordelt på 200 lande.

Google+ er det næststørste sociale netværk baseret på deres 540 millioner månedlige aktive brugere, men omkring halvdelen af disse besøger end ikke det sociale netværk skriver New York Times[68]. Det høje antal af brugere kommer

³14.774.400.000

⁴57 millioner twitter-profiler ud af 319 millioner indbyggere

⁵2.160.000.000

af, at Google kræver en Google+ profil, hvis man ønsker at anvende deres andre tjenester, som eksempelvis YouTube.

Instagram er et af de sociale netværk, hvor der bliver delt flest billeder og videoer. I følge data fra Instagram[41], har de månedligt over 200 millioner aktive brugere. Der bliver på månedlig basis delt ca. 20 milliarder billeder og videoer, hvilket er ca. 60 millioner om dagen. 65% af det totale antal brugere kommer fra lande udenfor USA.

2.5 Evaluering af sociale medier

De sociale medier har ændret både den måde vi kommunikerer med hinanden på, og den måde vi søger og indhenter informationer på. I dag antages det, at en stor del af kommunikationen blandt unge sker på de sociale medier, herunder Facebook. Det er længe siden de analoge medier som breve etc. blev erstattet med telefoni, som nu antageligvis vil blive erstattet af sociale platforme såsom Facebook og Twitter. De sociale medier har skabt endnu en kommunikationskanal, og denne kanal anvendes blandt andet til at indhente endnu flere informationer omkring brugerne og deres indbyrdes interaktion imellem. Disse data vil derfor være mulige at udnytte i forbindelse med dette projekt.

Effekten af de sociale medier har vist sig især på to områder. Den første, er måden hvorpå nyheder skabes og deles er flyttet fra det traditionelle medie til det sociale, hvor især *breaking news* bliver delt. Den anden, er hvordan de sociale medier gør det muligt for personer at hjælpe hinanden, når Google er tom for resultater.

Analysen af de større sociale netværk har givet et godt overblik over de mange forskellige typer netværk, der er til rådighed. Analysen viser, at der findes flere forskellige typer sociale netværk med hver deres fordele og ulemper, der i forbindelse med dette projekt kan anvendes.

Projektet er baseret på data, som er til rådighed på de sociale netværk. Det er derfor essentielt, at der er den nødvendige mængde data, således at projektet kan blive en succes. Med udgangspunkt i de største sociale netværk blev der udarbejdet en oversigt over hvor meget, de hver især gør tilgængelige. Denne viste, at der er den tilstrækkelige mængde data til rådighed og at en stor mængde af det totale antal profiler faktisk er aktive på månedlig basis. Her bidrager de aktivt ved at dele store mængder af hhv. opdateringer, billeder og/eller videoer.

Markedsanalyse

For at få succes på det hurtigt bevægende digitale marked er det vigtigt at opnå viden omkring eksisterende løsninger og finde muligheder i differentiering, før selv at påbegynde udvikling af et nyt produkt. Det følgende kapitel er en analyse af markedet, et udsnit af de vigtigste konkurrenter og en evaluering af deres produkter i forhold til at skabe et overblik over, hvordan et nyt produkt vil kunne differentiere sig fra de eksisterende løsninger.

3.1 Det digitale marked

Der findes i dag forskellige markeder for digitale produkter, og i forbindelse med dette projekt vil der være lagt fokus på markedet for SaaS¹-produkter. Tidligere har virksomheder været forpligtet til at købe eller udvikle samt vedligeholde deres egen software, men SaaS-plattformen har givet virksomheder et reelt alternativ. Fremfor at vedligeholde eget software, så gør SaaS-plattformen det muligt for virksomheder at kunne abonnere på online-tjenester, hvor software tilbydes i *skyen*², hvilket helt fjerner begrænsninger og gør det muligt at eksekvere applikationer uafhængigt af klient-plattformen.

En af de større SaaS-udbydere på markedet[61] beskriver følgende, som værende nogle af fordelene ved anvendelsen af SaaS.

- Nem tilgang via internet, hvor virksomheder kan tilgå deres software på alle tidspunkter af døgnet uanset platform.

¹Software-as-a-Service

²Online data ressourcer

- Baseret på abonnement og har som udgangspunkt, ingen nødvendige opstartsomkostninger.
- Skalerbare produkter, således at produktet matcher virksomhedernes forbrug.

3.1.1 Forretningsmodeller for digitale produkter

Overordnet set findes der på markedet to typer forretningsmodeller, **freemium** og en **betalingsløsning med prøveperiode**. Med udgangspunkt i de fundne konkurrenter i afsnit 3.2 er de to typer af forretningsmodeller analyseret og herunder beskrevet.

3.1.1.1 Freemium

Freemium er ikke en ny forretningsmodel og anvendes på forskellige markeder, herunder markedet for mobilapplikationer. Forretningsmodellen er bygget op omkring eksempelvis to typer af det samme produkt, hvor det ene tilbydes gratis og det andet i en betalingsudgave.

Det positive ved anvendelsen af denne forretningsmodel er, at den er yderst effektiv blandt startup-virksomheder, hvor et produkt ønskes udbredt til en stor skare af brugere, hurtigt. Freemium giver mulighed for, at brugere kan oprette sig gratis og uden nogen form for tidsrestriktioner afprøve og teste produktet for på den måde at finde ud af, om produktet passer til deres behov.

Dropbox[21] er et eksempel på anvendelsen af freemium-forretningsmodellen. Her tilbydes 2gb gratis plads i *skyen*, der giver mulighed for at afprøve produktet. Den gratis plads er tilstrækkelig for at teste produktet, men ønskes der fuld udnyttelse af produktet, så kræver det opgradering til en betalingsløsning. Ved at tilbyde en gratis version af produktet, opbygger Dropbox deres kundebase i håb om senere at kunne tilbyde dem et betalingsprodukt.

Ikke alle cases er så positive som Dropbox. En stor del af mindre startup-virksomheder tilbyder deres produkter eller services gratis, i håb om at blive lige så succesfulde som Dropbox. Freemium strategien kan være svær at håndtere, og i artiklen fra Wall Street Journal[43] nævnes virksomheden Chargify, som et eksempel på, hvor galt det kan gå med anvendelsen af freemium-forretningsmodellen. Grænsen mellem gratis- og betalingsversionen var sat for

højt, hvilket resulterede i, at de havde langt færre betalende kunder og deraf ringe indtægt. I sidste ende måtte man ændre forretningsmodellen, og konvertere gratis-kunder til betalende kunder, for at undgå en konkursbegæring.

3.1.1.2 Betalingsløsning med prøveperiode

Ligesom anvendelsen af freemium-forretningsmodellen, så ses anvendelsen af en betalingsløsning også blandt de andre interessenter på markedet. Her tilbydes som udgangspunkt forskellige niveauer af produkter, hvor pris og funktionalitet går hånd-i-hånd.

Som Dan Counsell[10] nævner, er denne forretningsmodel særlig god, hvis der er tale om en veletableret virksomhed, der har en større mængde loyale kunder, som er klar til at købe ved en eventuel produktlancering.

Hvor man ved en freemium-forretningsmodel har et begrænset gratis produkt, så ses det ofte i en betalings-forretningsmodel, at man tilbyder det fulde produkt i en begrænset prøveperiode. Anvendelsen af en prøveperiode tillader, som tidligere nævnt i freemium, at mulige kunder kan prøve produktet, og efter endt prøveperiode købe produktet. Denne metode kan betegnes som *first fix is free*, hvor man via prøveperioden forsøger at gøre kunderne *afhængige* - produktet skal give dem lyst til mere efter endt prøveperiode. En udvidet version, som set hos Spotify[63] og Netflix[51], hvor der skal indtastes kreditoplysninger for at få en gratis prøveperiode. Her er de første 30 dage gratis, hvor man efter perioden bliver flyttet til en betalt løsning, medmindre man i perioden aktivt har opsagt abonnementet. Typisk er det gratis at opsige uden nogen form for afgift.

Det er primært større og velansete virksomheder, som kan tillade sig at benytte denne type forretningsmodel. På baggrund af et velkendt omdømme antages det, at kundesegmentet er langt mere professionelt og feature orienteret.

3.2 Konkurrenter

Det er vigtigt med et godt kendskab til eventuelle konkurrenter, når der skal udarbejdes et nyt produkt. På markedet for enterprise social media management applikationer, findes der en lang række konkurrenter. Følgende er en gennemgang af de der anses som de største og vigtigste i forhold til et eventuelt nyt produkt.

Overskrift.dk[56] Prisstruktur: Betalingsløsning med prøveperiode Giver

mulighed for at skabe et simpelt overblik over danske sociale medier. Interaktionen er relativ simpel, hvor valgte søgetermer anvendes til at indhente data fra forskellige sociale platforme. Et diagram illustrerer hvor mange gange søgetermerne bliver nævnt pr. dag, og det er muligt at se hver opdatering de er nævnt i. Nem opsætning, men også et begrænset udbytte.

HootSuite[36] Prisstruktur: Freemium Tilbyder både at holde øje med brands ud fra statistiske diagrammer fra diverse sociale netværk, men også mulighed for at poste på dem. For at benytte disse features, skal brugerne selv vælge søgeord og tilpasse opsætningen. Dette giver mulighed for et godt omfattende produkt, hvis det sættes korrekt op.

Falcon Social[28] Prisstruktur: Betalingsløsning med prøveperiode Minder meget om HootSuite, ved at give en bred oversigt over social interaktion på specifikke søgeord. Ved selv at skulle vælge søgetermer samt tilpasse opsætningen, kræver det tid og ressourcer for at få et godt udbytte.

Xplenty[89] Prisstruktur: Betalingsløsning med prøveperiode Tilbyder en specifik analyse af relevante samtaler på sociale netværk. Xplenty går et skridt længere med kundeinteraktionen, og kan også lytte på samtaler. Mulighed for integration med andre platforme. Kræver it-kendskab for opsætning.

Geckoboard[31] Prisstruktur: Betalingsløsning Tilbyder et bredt statistisk overblik, blandt andet over sociale medier. Da det ikke kun er rettet mod sociale netværk, så giver de et mere generelt overblik over indhentet data.

Brandwatch[7] Prisstruktur: Betalingsløsning Tilbyder en dybere analyse end andre på markedet. De tilbyder både stemningsanalyse, filtrering, kategorisering, emne beskrivelse og håndtering af op mod 25 sprog. Brandwatch kræver også at kunden sætter sig ind i hvordan og hvilken data de vil modtage. Prismæssigt ligger de højere end de ovenstående konkurrenter, men tilbyder også mere end dem.

3.3 Evaluering af markedsanalyse

Gennem en analyse af de forskellige interessenter på markedet er det fundet, at der overordnet set anvendes to forskellige forretningsmodeller. Begge forretningsmodeller har fordele og ulemper for en startup-virksomhed på et marked med mange interessenter.

Fordelene ved at anvende en eventuel freemium-forretningsmodel er, at man hurtigt kan komme på markedet og skabe en god kundebase, som senere hen kan bruges til at konvertere eksisterende gratis kunder til betalende kunder. Ulempen er, at et forkert valg af gratis features, kan være enden på projektet. En betalt forretningsmodel ses også anvendt på markedet, men her er det nødvendigt med en vis mængde interesserede kunder eller et velkendt brand.

I analysen for de største og vigtigste konkurrenter i afsnit 3.2 er det fundet, at de opfylder forskellige behov hos deres kunder.

Nogle tilbyder et simpelt og let overblik over, hvad der på de sociale medier bliver skrevet. Disse kræver ikke meget fra kundens side, og er både nemme og overskueligt at benytte. Andre giver en dybere analyse, og dermed et mere specifikt billede af den sociale omtale. Dette kræver samtidig, at kunden sætter sig ind i hvordan det pågældende produkt virker. De fleste tilbyder support til opsætning, men fælles for dem alle er, at kunderne skal bruge tid og ressourcer på at sætte sig ind i hvordan produktet anvendes, og hvordan det vil kunne benyttes for dem. Hvad der kendetegner de fleste konkurrenter er, at de viser og illustrerer data, uden videre tolkning eller forståelse af denne.

Brandwatch skiller sig ud ved at gå skridtet videre og tilbyde mere. Fra blot at vise data på forskellige måder, tilbyder de også stemningsanalyse og emne genkendelse, eksempelvis sport. De fortolker det indhentede data, lader data fortælle en historie, og skaber derfor mere værdi end ved blot at illustrere data.

Sammenfattende for konkurrenterne er, at hvis mere relevant information ønskes ud fra social data, så skal man selv opsætte eventuelle filtre manuelt, hvilket kan være en besværlig og vanskelig process. Hvis disse filtre ikke bliver defineret korrekt, vil det afspejle sig i de indhentede informationer.

Forretningsmodel

Dette projekt vil blive gennemført som et lean-startup projekt, hvor anvendelsen af fundamentale principper vil danne grundlag for et udarbejdet produkt. For at kunne indtage markedsandele på et hurtigt bevægende marked, skal produktet have den rette forretningsmodel. Det følgende kapitel vil beskrive hvorfor lean-startup er valgt, anvendte principper fra lean, samt udarbejdelsen af en lean forretningsmodel.

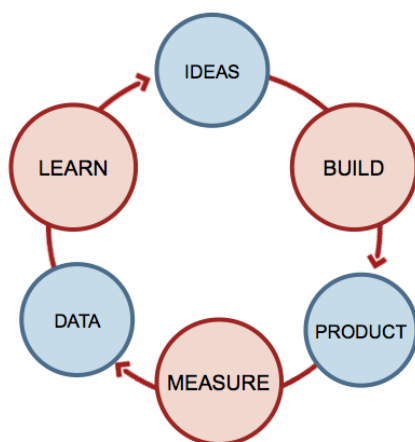
4.1 Hvorfor lean startup?

Processen for en lean startup kan bedst beskrives som en systematisk gennemgang af forretningsmodellen samt dets aktivitet. Gennem en iterativ proces vil disse systematisk blive evalueret. Formålet med denne systematiske gennemgang er, at man gennem iterationer går fra en plan A, der som udgangspunkt er mangelfuld, til en plan der faktisk virker, som man kan gå til marked med, som A. Maurya beskriver i bogen[49].

4.1.1 Build - Measure - Learn

En af de centrale komponenter i lean startup er build-measure-learn[65] modellen, der er illustreret i figur 4.1. Modellen er bygget op omkring tre faser *build*, *measure* og *learn*, som skal sikre konstant innovation.

Den første fase, *build*, beskriver det step, hvor man fra ideer og problemstillinger udarbejder et Minimum Viable Product[8]. Et Minimum Viable Product, herefter omtalt som MVP, er den version af produktet, som indeholder de mest



Figur 4.1: Konstant innovation ved Lean Startup, kilde: <http://lean.st/principles/build-measure-learn>

essentielle funktioner, som tillader hurtig indhentning og dermed også feedback på brugeres interaktion.

Den anden fase, *measure*, beskriver det efterfølgende step, hvor man ud fra det indhentede data måler effekten af elementerne i MVP'et. Det indhentede data vil da ligge til grund for den næste fase.

Den sidste fase, *learn*, beskriver det step, hvor man på baggrund af det indhentede data validerer produktet. På baggrund af denne analyse vil man her skulle bestemme, hvorvidt produktet eksempelvis er efterspurgt, om det skal redefineres eller helt droppes.

4.2 Lean forretningsmodel

En generel forretningsplan henvender sig til projekter, hvor der er mere veldefinerede grænser, eksisterende krav til en løsning og et prædefineret produkt. Dette projekt vil blive gennemført som et lean-startup projekt, og der findes derfor intet klart billede af løsningen, det færdige produkt eller dets forretningsmodel, og hvorvidt det faktisk er efterspurgt. Anvendelsen af lean kan gennem den iterative proces eliminere de ukendte faktorer, således at man step-for-step gennem processen evaluerer på, hvorvidt man skal gå videre eller ej. På denne måde, at kan man øge sandsynligheden for succes og dermed også undgå at

bruge unødvendige ressourcer. For at illustrere dette, vil der i projektet blive udarbejdet en lean canvas model, og dette er nærmere beskrevet i afsnit 4.2.1.

4.2.1 Lean Canvas Model

Anvendelsen af en lean canvas model gør det muligt at beskrive samtlige aspekter af produktet i et enkelt diagram. A. Maurya beskriver i bogen[49][s. 23], at netop dette canvas er det rette format for hurtigt at udarbejde forskellige forretningsmodeller til produktet, prioritering om hvor man bør starte, og især at man løbende ikke blot bliver klogere, men også benytter den nye viden til at skabe en bedre forretningsmodel for produktet.

Forretningsmodellen er bygget op omkring ni segmenter, som alle vedrører forskellige aspekter af produktet. Følgende er en liste over de segmenter samt rækkefølge, som forretningsmodellen indeholder.

- Kundesegment/Første kunder
- Problem samt eksisterende alternativer
- Unique Value Proposition
- Løsninger
- Kanaler
- Indtægtskilder
- Prisstruktur for udgifter
- Nøgleaktiviteter
- Urimelige fordele

Undervejs gennem projektet vil forretningsmodellen blive evalueret og ændret, hvis nødvendigt. Disse iterationer vil gøre det muligt at gå fra en "*Plan A*" til en forretningsmodel, som faktisk virker, som også A. Maurya beskriver i bogen[49].

Den første forretningsmodel er illustreret i figur 4.2, og i de følgende afsnit vil hvert segment af forretningsmodellen blive yderligere beskrevet.

Tagster - Forretningsmodel - første iteration

PROBLEM At indhente information omkring et specifikt emne At indhente kun relevant data og få frasorteret unødvendig data At indhente holdninger og meninger ud fra et brand EXISTING ALTERNATIVES BrandWatch Overskrift.dk Twitter HootSuite Falcon Social	SOLUTION Et emne-filter, som kan indhente data ud fra et emne Et spam-filter, som kan fjerne al ikke relevant data En metode, således meninger og holdninger kan udtrækkes	UNIQUE VALUE PROPOSITION Gør hvad I gør bedst, og lad os fortælle hvad folk mener om Jer Intet behov for søgeord eller svære forespørgsler, det hele sker automatisk Vi gør social data spiseligt	UNFAIR ADVANTAGE	CUSTOMER SEGMENTS #1 Mindre webshops #2 Marketingbureauer #3 Nyhedsbureauer EARLY ADOPTERS Sager indblik i social medie omtale Ønsker en dybere forståelse af hvad folk mener om specifikke emner
COST STRUCTURE Hosting omk. På nuværende tidspunkt gratis lokal server. Udviklings timer er på nuværende tidspunkt gratis. Forventes at tilbyde partnerskab til eventuelle nye udviklere. AdWords kampagner. Afhænger af CTR. Budget sat lavt til 20kr/dag. Fremtidig hosting hos Google Cloud: Mindst kr. 212,39/md. Variabel omk. forbundet med mængde af data indhentet	KEY METRICS Hovedområde: Indhente, analysere/filtrere og præsentere social data Succeskriterie: Opbygge en virksomhed med 10 faste kunder efter første år	HIGH-LEVEL CONCEPT Social data analyse uden al det unødvendige Social data analyse gjort simpelt	CHANNELS Venner og bekendte Mund-til-mund metoden Sociale medier (viral markedsføring) herunder Facebook og Twitter Landing-page AdWords kampagner	
		REVENUE STREAMS #1 "Basis" med 7-dages prøveperiode uden binding, herefter 499kr/md. #2 "Godt klædt på" med 14-dages prøveperiode uden binding, herefter 999kr/md. #3 "Overdrevet" med 30-dages prøveperiode uden binding, herefter 1999kr/md.		

Lean Canvas is adapted from The Business Model Canvas (BusinessModelGeneration.com) and is licensed under the Creative Commons Attribution-Share Alike 3.0 Un-ported License.

Figur 4.2: Forretningsmodel for første iteration

4.2.1.1 Kundesegment / Første kunder

Produktet i dette projekt er baseret på et bestemt kundesegment, således at man kommer ud til de helt rette kunder. Dette segment har fokus på kundesegmentet samt hvilke typer af kunder produktet henvender sig til.

Ud fra en brainstorm blev de overordnede kundesegmenter defineret og følgende er en liste over disse.

- Mindre virksomheder herunder webshops
- Marketingbureauer, som sælger løsningen til deres kunder (politik, events, virksomheder)
- Nyhedsbureauer (Radio 24/7, BT¹, EB² eller researchere)
- Større virksomheder med mange kunder (service/support)
- Public service herunder tv (DR³, TV2)

¹Berlingske Tidende

²Ekstra Bladet

³Danmarks Radio

Nogle af disse segmenter søger en forståelse for hvad personer mener om et givent produkt eller service. Nyhedsbureauer søger en dybere research på populære emner og ønsker at udforske, hvad personer snakker om ud fra disse. De har dog alle det til fælles, at de søger indblik i, hvad personer bidrager med og deres holdning til et givent emne. Dette vil også være kendetegnende for de første kunder.

Der er ud fra listen valgt tre kernekunder, der er tilføjet i segmentet i forretningsmodellen, hvilket er illustreret i figur 4.2.

4.2.1.2 Problem samt eksisterende alternativer

I dette segment er der fokus på at identificere kundernes eksisterende problemstillinger, samt eventuelle eksisterende løsninger, som de anvender.

I forbindelse med at udarbejde en liste over problemstillinger ud fra kundens synspunkt, er der først fundet en overordnet problemstilling og herefter anvendt 5-why's-metoden[83], således de egentlige problemstillinger kan blive fundet.

Den overordnede problemstilling er, at **det kan være svært, at få et overblik over emner skrevet på sociale medier**. Følgende er en liste over de problemstillinger, der blev fundet i forbindelse med anvendelsen af 5-why's-metoden.

1. At trække viden ud af sociale medier
2. At indhente kun relevant data og få frasorteret unødvendig data
3. At indhente information omkring et specifikt emne
4. At drage konklusion ud fra indhentet data
5. At indhente holdninger og meninger ud fra et brand

I kapitel 3 er der blevet udarbejdet en markedsundersøgelse, og ud fra de eksisterende interessenter på markedet er det fundet, at følgende interessenter vil kunne afhjælpe eller helt løse problemstillingerne fundet.

- BrandWatch
- Overskrift.dk
- Twitter

- HootSuite
- Falcon Social

De tre mest sandsynlige problemstillinger og eksisterende løsninger er blevet valgt, og er herefter blevet tilføjet til segmentet i forretningsmodellen i figur 4.2.

4.2.1.3 Unique Value Proposition

Et unique value proposition, herefter omtalt som UVP, definerer netop det som tilbydes gennem produktet, men også hvad der gør det unikt. Målet med denne er, at kunne konvertere mulige uvidende kunder til faktiske kunder, men vigtigst er ikke at sælge, men at fange deres opmærksomhed, som også A. Maurya beskriver i bogen[49][s. 29].

For at finde eksempler på gode UVP'er, er der taget udgangspunkt i den vigtigste problemstilling fra problem-segmentet i afsnit 4.2.1.2. Den vigtigste problemstilling er **at indhente information omkring et specifikt emne**, og baseret på denne er følgende UVP'er blevet udarbejdet.

- Gør hvad I gør bedst, og lad os fortælle hvad folk mener om Jer
- Intet behov for søgeord eller svære forespørgsler, det hele sker automatisk
- Vi gør social data spiselig

Ved at benytte high-level koncept-metoden[87] er det muligt at udarbejde en one-liner, som beskriver produktet på en abstrakt måde, således at alle nemt og simpelt vil kunne forstå hvad produktet omhandler. Dette gør det også nemmere for folk at dele budskabet om produktet, som er videre beskrevet i kanal-segmentet i afsnit 4.2.1.5. Følgende er to mulige high-level koncepter, som begge beskriver produktet på en abstrakt måde, uden at gå i detaljer.

- Social data analyse uden al det unødvendige
- Social data analyse gjort simpelt

De udarbejdede UVP'er samt high-level koncepter er illustreret i figur 4.2.

4.2.1.4 Løsninger

I dette segment vil der søges efter løsninger på de problemstillinger, der er skitseret i afsnit 4.2.1.2. Målet er at finde løsninger, der sammen med UVP'erne i afsnit 4.2.1.3 kan danne rammerne for et MVP⁴, som vil blive udarbejdet senere i projektet. Følgende er en liste over mulige løsninger på de nuværende problemstillinger.

- Et emne-filter, som kan indhente data ud fra et emne
- Et spam-filter, som kan fjerne alle ikke relevante data
- En metode, hvorpå meninger og holdninger kan udtrækkes

Som A. Maurya beskriver i bogen[49][s. 32], er det godt at holde disse løsninger simple og ikke gå for meget i detaljer. Dette sker på baggrund af, at man ofte kan se sig blind på detaljen i en given løsning, og dermed ikke fokusere på faktisk at løse problemstillingen. De fundne løsninger er illustreret i figur 4.2.

4.2.1.5 Kanaler

Gennem kanalerne findes der nye mulige kunder, og ud fra kundesegmentet i afsnit 4.2.1.1, er det muligt at finde de rette kommunikationsveje, således at man bedre kan nå ud til kunderne. Kanalerne er blevet defineret ud fra de kunder, som **søger indblik i social medie omtale**, eller **ønsker en dybere forståelse af, hvad folk mener om specifikke emner**. Følgende er de kanaler, som ønskes anvendt til at nå ud til de potentielle kunder.

- Venner og bekendte
- "Mund-til-mund"metoden
- Sociale medier (viral markedsføring), herunder anvendelsen af Facebook og Twitter
- Landing-page
- AdWords kampagner

⁴Minimum Viable Product

Først og fremmest vil der blive anvendt "mund-til-mund metoden", hvor man via venner og bekendte spreder budskabet om produktet. Her kan med fordel anvendes både de fundne UVP'er fra afsnit 4.2.1.3, men også i særdeleshed de tilhørende high-level koncepter. Som A. Maurya beskriver i bogen[49][s. 33], så er at fejle i kommunikationskanalerne en af de ting, som gør at startup virksomheder fejler. Allerede fra dag 1 i projektførløbet, vil sociale medier herunder Twitter blive anvendt, hvorfor en twitter-profil⁵ er blevet oprettet til formålet. Gennem denne profil vil både spyshots og teasers på produkter blive delt. Ønsket er her at få feedback fra de følgere, som undervejs i projektet vælger at følge twitter-profilen. En Facebook-side samt AdWords-kampagner vil senere kunne blive anvendt som kanaler til kommunikation, men på nuværende tidspunkt er Twitter tilstrækkeligt i forbindelse med dette projekt.

Kanalerne skal sørge for at drive trafik til en opsat landing-page, hvor produktet vil blive beskrevet i detaljer. De valgte kanaler er illustreret i figur 4.2.

4.2.1.6 Indtægtskilder

I dette segment beskrives indtægtskilder for produktet. For at finde markedsprisen, er der taget udgangspunkt i de eksisterende interessenter i afsnit 3.2, og hvordan de har prisfastsat deres produkter. Ud fra deres priser og features vil det være nemmere at vurdere, hvordan produktet skal prisfastsættes.

I afsnit 3.1.1 blev markedet undersøgt for at finde de generelt anvendte prisstrukturer. Et mix af freemium og betalingsløsning med prøveperiode blev fundet. For primært udelukkende at tiltrække mere seriøse kunder vælges en betalingsløsning, hvor der vil blive tilbudt en prøveperiode. Dette understreges af Liam Gooding, CEO af *trak.io*, som beskriver betalende kunder, som værende mere værdifulde jf. produktudviklingen.

„Free Users Help You To Build A Cool Product. Paying Customers Help You To Develop A Valuable Product“ [69], Gooding (2014).

For at matche både de forskellige kundesegmenter og dermed også kundernes enkelte behov, vil der tilbydes forskellige udgaver af samme produkt. Tre produkter er valgt, hvor pris og features går hånd-i-hånd. Ved at tilbyde tre forskellige udgaver, kan kunderne frit vælge efter behov, men også efter budget. På denne måde tilstræbes det at kunne imødekomme kundernes individuelle behov. I det følgende beskrives de tre forskellige produkter samt deres prisstruktur. Det er på nuværende tidspunkt for tidligt i processen at bestemme, hvilke features og

⁵@Tagsterdk findes på <https://twitter.com/tagsterdk>

antallet af dem, som vil være inkluderet i de forskellige udgaver. Begrænsninger på antal søgetermer, begrænset historik er begge features, som kunne danne grundlag for eventuelle begrænsninger i de tre udgaver af produktet.

- "Basis", som inkluderer 7-dages prøveperiode uden binding
- "Godt klædt på", som inkluderer 14-dages prøveperiode uden binding
- "Overdrevet", som inkluderer 30-dages prøveperiode uden binding

Ved at tilbyde en gratis prøveperiode i samtlige produkter, er det muligt for kunderne at afprøve produktet. Formålet med at tilbyde en gratis prøveversion er, at give kunderne de værktøjer, som løser deres nuværende problemer, og som det også er beskrevet i afsnit 3.1.1.2 at gøre kunderne "afhængige" og lyst til mere efter endt prøveperiode. Ved ikke at tilbyde nogen form for binding eller registrering af kreditkort oplysninger, gøres det simpelt at anskaffe sig en gratis prøveperiode. Inden endt prøveperiode vil kunden enten modtage en email eller en påmindelse om, at produktet snart vil udløbe og eventuelt tilbyde dem en rabat den første måned, hvis de registrerer deres første betaling.

Produktnavne som *basis*, *godt klædt på*, og *overdrevet* er bevidst valgte. Navnene fortæller, helt jordnært, hvad de indeholder, hvor det mange gange kan være svært at gennemskue, hvem et Pro eller Enterprise produkt henvender sig til. En mindre virksomhed vil kun have basis-behov, og vil alt andet lige vælge basis-produktet frem for de to andre.

De tre forskellige produkter er illustreret i figur 4.2.

4.2.1.7 Prisstruktur for udgifter

Der er både faste og variable omkostninger forbundet med at holde et system kørende og få udviklet alle nødvendige funktioner. Følgende er derfor en gennemgang af de omkostninger, som er forbundet med dette projekt, samt et overslag på eventuelle fremtidige udgifter.

Dette projekt indeholder en web-applikation, som på nuværende tidspunkt vil blive kørt på en lokal server, hvor strøm er den eneste udgift forbundet hermed. Denne udgift er dog så lille, at den ingen indflydelse har på projektet, og vil derfor ikke medtages. De nødvendige funktioner, samt udviklingstimer forbundet med udarbejdelsen af et MVP, bliver ej heller medregnet, da timerne stilles gratis til rådighed i løbet af dette projekt.

Når projektfællobet er afsluttet, vil produktet blive flyttet ud i Google Cloud, således at ressourcer fremadrettet kan sættes efter behov. Baseret på de mest nødvendige ressourcer såsom en web og data mining server, samt en database til at lagre data i. Mindsteprisen for dette er kr. 212,39 og udregningen er illustreret i figur 1 i Bilag A. Prisen afhænger direkte af, hvor populært produktet bliver, og udgiften forbundet med plads i skyen er derfor en variabel udgift med en mindste pris. Der er taget højde for, at mængder af indhentede data i forbindelse med konto i prøveperiode, også vil blive lagret i databasen. Hvis kunden ikke ønsker at bibeholde en konto, kan unødvendige data slettes, og på den måde kunne mindske dataforbrug.

Hvis udviklingstimer også fremover ikke medregnes som en udgift, skal der blot være en enkelt kunde på et basis produkt, for at produktet kan løbe rundt. De nuværende og fremtidige udgifter er illustreret i 4.2.

4.2.1.8 Nøgleaktiviteter

I dette segment beskrives hvordan nøgle aktiviteterne for produktet vil være. A. Maurya beskriver[49], at for startup virksomheder foreslås at fokusere primært på nøgleaktiviteter, som kunder kommer til at benytte samt succeskriterier. For at opnå succeskriterierne er det nødvendigt, at kundens problemstillinger afhjælpes ved løsninger som produktet tilbyder. På baggrund af dette, vil produktets hovedområde samt succeskriterie blive defineret.

Produktets hovedområde er at indhente, analysere og præsentere data på en måde, så det er nemt at arbejde med. Det er derfor essentielt for produktets succes, at metoder til indhentning og analysering af data er effektive, samt at data er præsentable.

Succeskriterier kan være svære at definere, især når omkostninger ved produktet ikke er særligt høje, som beskrevet i afsnit 4.2.1.7. Hvis arbejdskraften stadig forbliver kost nul, vil det være muligt at få produktet til at løbe rundt med en enkelt basis kunde. Den ene kunde gør det dog ikke rentabelt, og et succeskriterie på ti faste kunder efter et år, er derfor ikke for uambitiøst. Ti faste basis kunder vil bidrage med en månedlig indtægt på kr. 5.000 jf. prisstrukturen beskrevet i afsnit 4.2.1.6. Denne månedlige indtægt vil være et passende engagement til videreudvikling af produktet og give kunderne endnu mere for deres penge.

Det opstillede hovedområde og succeskriterie er illustreret i figur 4.2 .

4.2.1.9 Urimelige fordele

I dette segment beskrives hvilke urimelige fordele produktet ville skabe over for konkurrenterne. Som A. Maurya beskriver i bogen [49][s. 42], kan det være svært at definere eventuelle urimelige fordele for en startup. På nuværende tidspunkt vil projektet kun indeholde kode og features, og disse kan ikke betegnes som urimelige fordele, da disse nemt kan enten kopieres eller andre kan købe sig til. Et eksempel på en fordel vil i fremtiden kunne være en stor kundedatabase, eller et velanset og kendt brand.

Baseret på markedsanalysen i kapitel 3 er det fundet, at ingen eksisterende løsninger på en nem måde giver en dybere forståelse af indhentede data. Ud fra UVP beskrevet i afsnit 4.2.1.3, vil produktet kunne finde relevante data uden at kunden explicit skal tilføje søgetermer eller filtre. Dette vil give produktet en fordel på markedet, men kan ikke antages som værende en urimelig fordel, da begge metoder vil kunne kopieres med tiden.

Ved at lade dette segment stå tomt i figur 4.2, vil det stå som en konstant påmindelse om, hele tiden at stræbe efter funktioner, som vil kunne give produktet fordele på markedet.

4.3 Evaluering af forretningsmodel

Projektet bliver kørt som et lean-startup projekt, hvilket igennem en systematisk gennemgang af forretningsmodellen, at kunne evaluere gennem iterationer på forretningsmodellen, og dermed kunne gå fra "*Plan A*" til en plan man kan gå til marked med. Dette gør det muligt, at man undervejs kan eliminere ukendte faktorer og undersøge hvorvidt produktet faktisk er efterspurgt. Dette kan øge sandsynligheden for succes.

Ved at gøre brug af Lean Canvas Model, kan der dannes et overblik over essentielle segmenter i forbindelse med udarbejdelsen af produktet. Gennem løbende evaluering og ny viden, kan der skabes en mere præcis forretningsmodel.

For at komme ud til de rette kunder, baseres produktet på et bestemt kundesegment, hvor kunderne har det til fælles, at de alle søger indblik i hvad personer bidrager med, og deres holdning til et givent emne. Kundesegmentet er i dette projekt blevet defineret som mindre webshops, marketingbureauer og nyhedsbureauer.

Ud fra den overordnede problemstilling, hvor det kan være svært at få et overblik over emner skrevet på sociale medier, blev kundesegmentets eksisterende problemstillinger fundet. At trække viden ud af de sociale medier, frasortere det unødvendige data og herfra indhente information om et specifikt emne, blev fundet som de problemstillinger, der er vigtigst for kunderne i forhold til dette projekt. Baseret på problemstillingerne og markedsanalysen blev eksisterende løsninger på markedet fundet.

For at kunne promovere produktet bedst muligt, blev flere forskellige udgaver af et unique value proposition udarbejdet. Disse skulle sørge for ikke alene at sælge produktet, men at skabe opmærksomhed og interesse, således at de kan viderebringe budskabet om produktet. For nemmere at få potentielle kunder til at forstå produktet, er der udarbejdet to high-level koncepter, som på en abstrakt måde beskriver produktet.

Ud fra de tre vigtigste problemstillinger blev der udarbejdet mulige løsninger til disse. Løsningerne blev holdt på et relativt enkelt niveau uden at gå i detaljer med, hvordan de enkelte løsninger vil afhjælpe problemstillingerne. Løsningerne som produktet skal indeholde er emne-filter, et spam-filter og en metode, hvor meninger og holdninger kan udtrækkes.

Gennem de rette kommunikationskanaler skal produktet promoveres. I dette projekt er der taget udgangspunkt i kanaler, hvor potentielle kunder søger forståelse og indblik i specifikke emner indenfor sociale medier. Især anvendelsen af venner og bekendte via "mund-til-mund"metoden blev fundet relevant, men også anvendelsen af sociale medier herunder Facebook og Twitter. De valgte kanaler skal bistå med at henvise besøgende til en opsat landing-page.

Der er i dette projekt taget udgangspunkt i de eksisterende konkurrenters prisstruktur. Det blev fundet, at den bedste løsning for dette produkt ville være en betalingsløsning, da betalende kunder blev anset som værende mere værdifulde mht. videreudvikling af produktet. For at dække de forskellige kunders behov, blev valget på tre typer af produktet fastlagt. En eventuel prøveperiode vil give kunderne en *smagsprøve*, og forhåbentlig tilmelde sig som betalende kunder.

Det blev fundet at udgifter forbundet med produktet er proportionalt med antallet af kunder. På nuværende tidspunkt findes ingen essentielle udgifter, men når produktet går live, vil de få udgifter være dækket af få kunder.

For at produktet skal blive en succes er det nødvendigt, at kundernes problemstillinger bliver afhjulpet gennem de ydelser produktet tilbyder. For at kunne videreudvikle produktet og dermed give kunderne mest muligt for deres penge, er et succeskriterie på ti faste kunder indenfor det første år fundet nødvendigt.

Da produktet på nuværende tidspunkt ikke indeholder andet end kode samt nødvendige features kan produktet ikke anses for at ligge inde med urimelige fordele hvad angår de nuværende interessenter på markedet.

1. iteration

Denne første iteration vil forsøge at skabe overblik over det videre forløb for projektet. I dette kapitel vil det valgte sociale netværk blive defineret, samt udarbejdet hypotese på baggrund af forretningsmodellen, som skal ligge til grund for det videre forløb.

5.1 Twitter

Det blev i afsnit 2.3 beskrevet kendetegnene ved de største sociale netværk. Følgende vil være en gennemgang af disse, samt baggrunden for valg af medie.

Facebook er det mest udbredte sociale netværk, men på baggrund af privat/offentlig-problemstillingen hvor ikke alle opdateringer er frit tilgængelige. Sammen med den manglende funktion at kunne udforske en samtale ud fra et hashtag har gjort, at Facebook ikke vil blive anvendt i forbindelse med første version af produktet.

Ligeledes er LinkedIn her fravalgt pga. dets fokus på det forretningsmæssige, hvilket ikke er relevant for dette projekt - i hvertfald ikke i denne version.

Instagram har vist sig at være god til billeder, men pga. manglen på underbyggende tekst og overdrivelsen af hashtags er dette også fravalgt. Dog vil billederne senere hen kunne styrke indholdet af det indhentede relevante data.

Google+ er et niche netværk, og er fravalgt pga det manglende fokus fra mediernes side, da det er sjældent at se i nyhederne, at kommentarer etc. kan følges på deres Google+ profil.

Twitter har vist sig at være det bedste valg til dette projekt og det er sket på baggrund af følgende.

- Anvendes ofte som socialt medie under forskellige typer af events, herunder alt fra valgkampagner[72] til Eurovision[71] etc.
- Et alsidigt medie, som via hashtags gør det nemt at følge mange personers meninger og holdninger om et valgt emne.
- Data er offentligt tilgængelig og er modsat f.eks. Facebook's private netværk, langt nemmere at udtrække relevant information fra.

5.2 Opstilling af hypoteser

Der blev i forretningsmodellen i afsnit 4.2.1.4 opstillet løsninger til de fundne problemstillinger på markedet beskrevet i afsnit 4.2.1.2. Disse løsninger skal være indeholdt i et udarbejdet MVP. Inden metoder til løsningerne kan findes, er det nødvendigt at undersøge hvordan data kan indhentes. Derudover vil det blive undersøgt hvad social data ellers kan beskrive ud fra forskellige emner. På baggrund af dette er følgende hypoteser opstillet.

- Det antages, at der fra det sociale netværk Twitter kan indhentes struktureret data
- Det antages, at det ud fra det indhentede data er muligt, at kunne filtrere på specifikke emner, samt kunne frasortere ikke relevant data
- Det antages, at det ud fra filtreret data er muligt at kunne udtrække meninger og holdninger, samt kunne rangere og sortere på hhv. brugere og tweets
- Det antages, at der på baggrund af det sorterede og filtrerede data er muligt at udarbejde et MVP man kan gå til marked med
- Det antages, at der i social data kan findes mønstre ud fra forskellige begivenheder

I de følgende kapitler vil hypoteserne løbende blive testet, og gennem iterationer vil det vise sig, hvorvidt der skal foretages ændringer.

5.3 Evaluering af 1. iteration

I dette kapitel blev Twitter fundet som det foretrukne sociale netværk til projektet. Her blev der lagt vægt på åbenheden, tilgængeligheden samt anvendelsen under større begivenheder.

Ud fra løsninger til problemstillinger fundet i forretningsmodellen blev der opstillet hypoteser, der gennem dette projekt skal danne grundlag for produktet. Yderligere blev der defineret en hypotese for hvad social data ellers kan bidrage med, og hvad produktet ellers kan indeholde.

Gennem en iterativ proces vil hypoteserne blive forsøgt testet og yderligere analyseret.

KAPITEL 6

2. iteration

I anden iteration vil der gennem hypoteser blive testet, hvorvidt data fra Twitter kan indhentes. Der vil indledningsvis blive gået i dybden med at teste forbindelsen til Twitter, analyser af forskellige indsamlingsmetoder samt data struktur heraf.

6.1 Hypoteser

Inden videre analyse af hvordan data indhentes, er det nødvendigt at opstille hypoteser for, hvorvidt det faktisk er muligt at indhente data fra Twitter. Følgende hypoteser er derfor opstillet.

- Det antages at kunne indhente data fra Twitter
- Det antages, at data er struktureret på en sådan måde, at det kan anvendes til videre analyse

I følgende afsnit vil det forsøges at teste de opstillede hypoteser gennem forskellige indsamlingsmetoder.

6.2 Indsamling af data

Stort set al data på Twitter er offentlig tilgængeligt, som det også er beskrevet i kapitel 2. Twitter er et stærkt voksende medie i Danmark[3], dog foregår

størstedelen af alle tweets stadig i stigende grad på engelsk. For at gøre indsamlingen bedre, bredere og større, vil dette projekt derfor koncentrere sig om tweets på engelsk. Da Twitter er et åbent og lettilgængeligt medie, indeholder dette derfor en stor variation af data, herunder tweets omkring virksomheder og deres omdømme. Der kan indhentes en stor mængde viden om virksomheder, baseret på tweets samt interaktionen brugerne imellem. Dette både direkte fra virksomhedernes offentlige twitter-profiler og indirekte fra samtaler, der opstår via twitter-brugernes interne netværk.

Følgende afsnit vil være en gennemgang af, hvordan data fra Twitter kan indhentes ud fra forskellige indsamlingsmetoder.

6.3 Forbindelse til Twitter API

Der kan opnås forbindelse til offentlig tilgængelig data på Twitter vha. tre metoder, herunder *public streams* såsom **firehose** og **sample**, men også *REST API*[14] i form af **search**. De tre metoder, som stilles til rådighed af Twitter, er i følgende afsnit beskrevet og analyseret.

- Firehose - Fuld adgang til data (ikke gratis tilgængeligt)
- Search - Tweets frem til idag ud fra specifikke søgekriterier
- Sample - Begrænset adgang til data i form af live streaming (gratis tilgængeligt)

Ideelt set ville fri adgang til Twitter's data være at fortrække, men idet data på Twitter er en del af deres forretningsmodel, er adgang til data begrænset. Denne begrænsning kan eventuelt omgås ved eksempelvis en streaming-tjeneste (automatisk), som hele tiden er opdateret med seneste tweets, eller i form af en søgning (manuelt), som ud fra bestemte tidsintervaller henter seneste tweets.

Firehose giver fuld adgang til data, men dette fører en økonomisk udgift med sig. Løsningen giver adgang til både *public streams* og *REST API'er*. Alternativer til at få adgang til al data er eksempelvis at benytte en mellemmand, som har adgang til firehose og stiller data til rådighed. Gnip[32] og Datasift[11] er begge eksempler på dette og giver begge mulighed for at få firehose adgang. Hvor firehose er rå og ufiltreret data, så tilbyder både Gnip og Datasift en nemmere måde at indhente information ud af det rå data på. Under udarbejdelse af dette projekt har Twitter købt Gnip[48], hvilket blot understøtter værdien i udtrukket information fra social data.

Search er en metode der stilles til rådighed af Twitter, og gennem dette *REST API* kan der indhentes tweets frem til idag ud fra specifikke søgekriterier. Da denne metode stilles gratis til rådighed, så kommer den med visse restriktioner. En af begrænsningerne er *rate limits*[16], hvor antallet af forespørgsler til Twitter er begrænset. I en twitter applikation er det tilladt at foretage 450 forespørgsler pr. 15 min[15]. Der er på nuværende tidspunkt ingen grænse for, hvor mange applikationer man kan oprette, udelukkende pga den manuelle proces ved dette. Hver forespørgsel kan indeholde op til 100 tweets, hvilket gør det muligt at indhente op med 45.000 tweet pr. 15 min.

Sample APIet er en tjeneste, som returnerer en tilfældig stikprøve af de tweets, der kommer fra Twitter. Mængden af udleveret tweets ved denne løsning lader til at være faldet fra 2 til 1%, siden sidste versionsskift på API'et. Diskussionen om hvorvidt det er 1%[64] eller 2%[27] varierer. I mange tilfælde afhænger det af, hvor mange af disse tweets det er muligt at hente, processere samt lagre. Spørgsmålet kan så bestå i om denne stikprøve er tilstrækkelig. Ifølge forskere fra Arizona State University[78] kan denne stikprøve godt være repræsentativ nok til at kunne udlede signifikante resultater. Da artiklen ikke omhandler spam, herefter benævnt hvid støj, bliver der ikke yderligere gået i dybden med tilstedeværelsen af tilfældige tweets. Dette vil blive afdækket yderligere i afsnit 7.3. Ved at definere forskellige parametre i *sample*, er det derved muligt at kunne filtrere på hhv. sprog, land, specifikke brugere og termer. Dette gør det muligt at returnere kun det ønskede data, hvilket også resulterer i mere relevant og præcis data.

6.4 Analyse af indhentet data

Som nævnt i afsnit 6.3, findes der tre forskellige metoder til at indhente data fra Twitter. En returneret tweet er opbygget som en *enhed*, og indeholder derfor både et tweet-objekt og bruger-objekt. Disse enheder er opbygget i en særlig data struktur, der er nærmere beskrevet på Twitters egen udvikler-side[13]. Udover selve teksten i en tweet, indeholder denne *enhed* også informationer om, hvem der har skrevet den, hvor og hvornår den er skrevet, en liste af hashtags anvendt, referencer til billeder/video andre url-referencer, samt hvorvidt der er tale om en retweet¹ eller original-tweet. Dette gør det betydeligt nemmere, senere at foretage filtrering og sammenligning af tweets.

For at få fat i et stort datasæt bruges *Sample API*. Data vil være forskelligt, og antages, at være repræsentativt[78] for hvad der findes på Twitter generelt. I forbindelse med dette projekt er der opstillet en lokal server, hvor en

¹En tweet, som er delt af andre brugere

MySQL-database er oprettet til at lagre de returnerede tweets. Datamodellen for databasen er illustreret i 6.1.

Field	Type	Null	Key	Field	Type	Null	Key
id	varchar(100)	NO	PRI	id	varchar(100)	NO	PRI
user_id	varchar(100)	NO	MUL	name	varchar(100)	NO	
created_at	datetime	NO		screen_name	varchar(150)	NO	
text_raw	varchar(500)	YES		description	varchar(300)	NO	
text_wordlist	varchar(500)	YES		followers_count	int(11)	NO	
coordinates	varchar(100)	NO		friends_count	int(11)	NO	
city	varchar(100)	NO		listed_count	int(11)	NO	
country	varchar(100)	NO		favourites_count	int(11)	NO	
hashtag_list	varchar(500)	NO		created_at	datetime	NO	
url_list	varchar(500)	NO		statuses_count	int(11)	NO	
user_mention_list	varchar(500)	NO		time_zone	varchar(100)	NO	
source	varchar(100)	NO		lang	varchar(100)	NO	
retweet	tinyint(1)	NO		profile_image_url_https	varchar(500)	YES	
retweet_count	int(11)	NO		klout_score	int(11)	NO	
retweet_id	varchar(100)	NO		rank	int(11)	NO	
in_reply_to_status_id	varchar(100)	NO					
rank	int(11)	NO					

Figur 6.1: MySQL Database struktur

Efter blot syv dage, er der blevet indhentet ca. **3,4 millioner** tweets, som er skrevet af ca. **2,5 millioner** unikke brugere, hvilket er illustreret i figur 6.2. Herfra er data klar til at blive videre analyseret. Det vurderes, at et datasæt på denne størrelse er tilstrækkeligt for dette projekt, og indsamlingen af flere tweet er derfor sat på hold. At der på blot en uge er indhentet så mange tweets, understøtter blot mængden af tweets til rådighed opgivet i afsnit 2.4.

```
mysql> select count(1) from coreapp_tweet;
+-----+
| count(1) |
+-----+
| 3444542 |
+-----+
1 row in set (0.01 sec)
```

```
mysql> select count(1) from coreapp_user;
+-----+
| count(1) |
+-----+
| 2569351 |
+-----+
1 row in set (3 min 37.41 sec)
```

Figur 6.2: Optælling af data gemt i databasen

De indhentede 3,4 millioner² tweets er, som beskrevet tidligere, repræsentative og et udsnit af, hvad der er tilgængeligt på Twitter. I et udsnit af ti tilfældigt udtrukket tweets, som er illustreret i figur 6.3, ses det tydeligt, at der i forhold til dette projekt også findes mange irrelevante tweets.

```
-----
| text_raw |
-----
| I feel guilty because I've only be in the 5sosfam for a little over a year like end of 2012 beginning 2013 |
| RT @deanballantynes: There's like 5 good acts going to T, why is every fucker buzzing out there nut |
| &amp; it's a done deal . Oh well . Lol |
| kill all your friends. |
| @smoshantony boeh wll there be Google search fun #3 ? |
| RT @meganni_coloff: @meganiizer _ lly 437 love &lt;t;3 #cynthianna #diathia which one? lol |
| 3 people followed me // automatically checked by http://t.co/i1L8eJ2Rsb |
| Happy birthday ben! Miss ya kiddo! |
| I don't accuse cause I know. |
| Smh... you ain't shit... lol... &amp; I mean that from the bottom of my soul... |
-----
```

Figur 6.3: Ti tilfældigt udtrukket tweets fra dataset på 3,4 millioner

Her illustreres det, at *Sample API* ikke er tilstrækkeligt alene, idet der bliver frigivet meget data, hvor en del af denne ikke direkte kan anvendes i dette projekt. Derfor er yderligere filtrering af tweets nødvendig for at filtrere irrelevante tweets fra og dermed have relevant data tilbage.

²3,444,542 tweets

6.5 Evaluering af 2. iteration

En forundersøgelse har vist, at størstedelen af tweets skrives på engelsk, og for at kunne indhente mest mulig data, er engelske tweets derfor grundlaget for dette projekt. Mange brugere og kunder deler erfaringer og meninger på de sociale medier omkring virksomheder, herunder deres produkter samt eventuelle services. For at udtrække data fra Twitter findes der tre metoder, **firehose**, **search** og **sample**. I dette projekt er det valgt primært at anvende **sample**, hvor der gratis er en begrænset mængde data tilgængelig. En undersøgelse af mængden af data til rådighed ved sample har vist, at på syv dage er mere end 3,4 millioner tweets indhentet, og kan derfor kategoriseres som værende *Big Data*. Dette er set i lyset af, at sample kun giver adgang til ca. 2% af alle tweets. I forbindelse med en tilfældig stikprøve af det indhentet data har det vist, at der findes en stor del irrelevant data, hvilket derfor kræver en yderligere filtrering, således at kun relevant data står tilbage. Denne iteration har givet et overblik over muligheder for indhentning af data, hvordan data er struktureret, og dermed også givet indsigt i, hvordan denne proces kan optimeres videre.

6.6 Validering af forretningsmodel

Det er i denne iteration blevet valideret, at det er muligt at indhente data fra Twitter gennem forskellige indsamlingsmetoder. Det indhentede data viste sig yderligere at være i en sådan form, at det vil kunne filtreres og sorteres ud fra forskellige parametre til videre analyse.

Det er herved påvist, at det er muligt at skabe et datagrundlag til løsningerne beskrevet i forretningsmodellen i afsnit 4.2.1.4, samt at kunne påvise hovedområdet i nøgleaktiviteterne i afsnit 4.2.1.8. Forretningsmodellen samt de berørte segmenter er illustreret i figur 6.4.

Tagster - Forretningsmodel

<p>PROBLEM</p> <p>At indhente information omkring et specifikt emne</p> <p>At indhente kun relevant data og få frasorteret unødvendig data</p> <p>At indhente holdninger og meninger ud fra et brand</p> <p>EXISTING ALTERNATIVES</p> <p>BrandWatch</p> <p>Overskrift.dk</p> <p>Twitter</p> <p>HootSuite</p> <p>Falcon Social</p>	<p>SOLUTION</p> <p>Et emne-filter, som kan indhente data ud fra et emne</p> <p>Et spam-filter, som kan fjerne al ikke relevant data</p> <p>En metode, således meninger og holdninger kan udtrækkes</p>	<p>UNIQUE VALUE PROPOSITION</p> <p>Gør hvad I gør bedst, og lad os fortælle hvad folk mener om Jer</p> <p>Inet behov for søgeord eller svære forespørgsler, det hele sker automatisk</p> <p>Vi gør social data spiseligt</p>	<p>UNFAIR ADVANTAGE</p>	<p>CUSTOMER SEGMENTS</p> <p>#1 Mindre webshops</p> <p>#2 Marketingsbureauer</p> <p>#3 Nyhedsbureauer</p> <p>EARLY ADOPTERS</p> <p>Søger indblik i social medie omtale</p> <p>Ønsker en dybere forståelse af hvad folk mener om specifikke emner</p>
<p>COST STRUCTURE</p> <p>Hosting omk. På nuværende tidspunkt gratis lokal server.</p> <p>Udviklings timer er på nuværende tidspunkt gratis. Forventes at tilbyde partnerskab til eventuelle nye udviklere.</p> <p>AdWords kampagner. Afhænger af CTR. Budget sat lavt til 20kr/dag.</p> <p>Fremtidig hosting hos Google Cloud: Mindst kr. 212,39/md. Variabel omk. forbundet med mængde af data indhentet</p>	<p>REVENUE STREAMS</p> <p>#1 "Basis" med 7-dages prøveperiode uden binding, herefter 499kr/md.</p> <p>#2 "Godt klædt på" med 14-dages prøveperiode uden binding, herefter 999kr/md.</p> <p>#3 "Overdrevet" med 30-dages prøveperiode uden binding, herefter 1999kr/md.</p>			
<p>KEY METRICS</p> <p>Hovedområde: Indhente, analysere/filtrere og præsentere social data</p> <p>Succeskriterie: Opbygge en virksomhed med 10 faste kunder efter første år</p>	<p>HIGH-LEVEL CONCEPT</p> <p>Social data analyse uden al det unødvendige</p> <p>Social data analyse gjort simpelt</p>	<p>CHANNELS</p> <p>Venner og bekendte</p> <p>Mund-til-mund metoden</p> <p>Sociale medier (viral markedsføring) herunder Facebook og Twitter</p> <p>Landing-page</p> <p>AdWords kampagner</p>		

Lean Canvas is adapted from The Business Model Canvas (BusinessModelGeneration.com) and is licensed under the Creative Commons Attribution-Share Alike 3.0 Un-ported License.

Figur 6.4: Forretningsmodel for anden iteration

3. iteration

Struktureret data er blevet indhentet og er klar til videre analyse. En stor del af data tilgængeligt har vist sig at indeholde en mængde ikke relevante tweets. I denne iteration vil der gennem hypoteser blive testet hvorvidt data kan filtreres, således kun relevant data står tilbage.

7.1 Hypoteser

Inden videre analyse af data er det nødvendigt at opstille hypoteser for, hvorvidt det er muligt at filtrere uønsket data fra. Følgende hypoteser er derfor opstillet.

- Det antages at kunne filtrere tweets ud fra specifikke søgetermer
- Det antages, at det er muligt ud fra det indhentet data at kunne filtrere reklamer og lign fra i tweets

Disse hypoteser er essentielle for at kunne indhente information omkring virksomheder ud fra tweets. Følgende afsnit er derfor en gennemgang af, hvordan der filtreres på virksomheders brand, samt hvordan ikke relevant data frasorteres.

7.2 Brand-filtrering

For at indhente relevant data, er en række virksomheder blevet valgt og deres specifikke søgetermer defineret, hvilket også illustrerer hvordan eventuelle kunder vil være præsenteret. Disse virksomheder og deres specifikke søgetermer er

illustreret i figur 7.1. De valgte brands er udvalgt på baggrund af, hvordan de hver især er repræsenteret på de sociale medier, og at et eksisterende kendskab til disse allerede haves. En del tweets om disse brands udtrykker sig enten markant positivt eller negativt. Et eksempel på dette kunne være *Eurovision*, idet der er mange tweets om dette, og hver især bærer præg af brugerens holdning.

id	name	search_terms
001	Apple	apple,ipad,iphone
002	Samsung	samsung
003	macdonalds	mcdonalds,macdonalds,mcdonalds,macdonalds,bigmac
004	digitalocean	digitalocean,digital ocean
005	cloudmagic	cloudmagic,cloud magic
006	att	att
007	nokia	nokia
008	amazon	amazon
009	Eurovision	#joinus,#eurovision,eurovision

Figur 7.1: Udvalgte brands samt deres specifikke søgetermer

De specifikke søgetermer for hvert brand er manuelt udvalgt ved at udforske tweets om hvert brand på Twitter. Herefter er der anvendt frekvensfordeling¹ således de mest anvendte ord kan findes. Disse søgetermer giver mulighed for, at fremadrettet data er mere relevant og målrettet.

Til sammenligning med allerede indhentede data, som beskrevet i afsnit 6.4, blev der tilføjet en adskilt database, hvor tweets fra de udvalgte brands er indeholdt. En tilfældig stikprøve på 200 tweets for brandet Apple, genererer følgende wordcloud som er illustreret i figur 7.2.

Det ses, at det indhentede datasæt er mere målrettet jf. de specifikke opsatte søgetermer. Det indeholder dog stadig ikke relevante tweets og de hyppigst brugte ord er *follow*, *retweet*, *followback* og lignende. Disse tweets, som kun er hvid støj skal frasorteres. Dette er yderligere beskrevet i afsnit 7.3.

7.3 Frasortering af ikke relevant data

Der blev i afsnit 6.4 beskrevet, hvordan et datasæt på 3,4 millioner blev indhentet, og ud af dem blev det fundet, at en stor del ikke var relevante. De ikke relevante tweets kan bedre beskrives som værende *hvid støj*, som ikke bidrager med at beskrive det pågældende brand.

I et nyt datasæt er der, på baggrund af de opstillede søgetermer, blevet indhentet ca. 750.000 nye brand specifikke tweets. Dette datasæt er dog ikke foruden

¹*frequency distribution*[86]



Figur 7.2: Wordcloud for 200 tweets til brandet Apple

hvid støj, hvilket kan skyldes de valgte søgetermer. Det ses nemlig, at populære termer anvendes i tweets, som kan kategoriseres som værende hvid støj, altså spam-relaterede tweets[88]. Twitter gør selv et stort stykke arbejde for at fjerne den hvide støj. Deres egen definition af spam ser således ud og er yderligere defineret på deres side[74].

„Posting harmful links to phishing or malware sites, repeatedly posting duplicate tweets, and aggressively following and un-following accounts to attract attention are just a few examples of spam on Twitter“ [70], Twitter Inc. (2010).

For yderligere at undgå hvid støj og misbrug har Twitter defineret en øvre grænse for, hvad der kan kategoriseres som normalt brug - alt over denne grænse vil derfor blive tolket som misbrug. Denne grænse, som også IAG.ME [39] beskriver, er sat relativt højt med et max antal tweets pr. dag til 1.000, samt 250 beskeder sendt direkte til andre brugere. Grundet den relativt høje grænse ses det i datasættet at brugere, som genererer store mængder af hvid støj, slipper gennem Twitters filter.

Der kan på baggrund af det indhentede data opstilles tre typer hvid støj, som ønskes frasorteret, idet de ikke bidrager positivt i datasættet i dette projekt. De tre typer af hvid støj er kategoriseret således.

- App-genererede tweets, herunder i særdeleshed spil
- Pornografisk materiale

- Tweets, hvor brugere i stor stil, er ude efter nye følgere

Et eksempel på de tre typer hvid støj er illustreret i figur 7.3. Det er antaget, at det ikke er de eneste tre typer hvid støj[62], som findes på Twitter, men i forbindelse med dette projekt, anses disse tre for værende de mest udbredte. Det er ofte set, at der bliver anvendt populære hashtags i denne type af tweets, og i figuren ses hashtags som *#iPhone* anvendt. Denne strategi kan betegnes som *hashjacking*[17], hvor man forsøger at overtage et hashtag, for på den måde at komme ud til flere brugere. Denne strategi anvendes også ofte i marketingkampagner[79], hvor produkter ønskes promoveret.

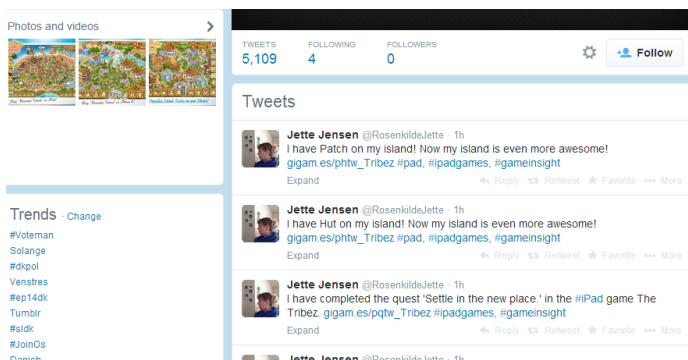
```

1 "I've collected 18,437 gold coins! http://t.co/KUzr754Nlh
2 \#iphone, \#iphonegames, \#gameinsight"\
3
4 "\#teen, \#pussy, \#movie, \#iphone, \#sex, \#porn Dick Addicted
5 Indian Chick Wins A Dick For A Splendid Fuck http://t.co/m7w8JQjB02"\
6
7 "RT @JJoinklamnder19: @Gain Followers"\

```

Figur 7.3: Tre former for hvid støj

Det er på figur 7.4 illustreret, hvorledes en bruger udsender app-genereret tweets. Der er her tale om en bruger, som gennem et spil har tilknyttet sin Twitter-konto. Det er med stor sandsynlighed, at brugeren ikke selv er aktiv på sin Twitter-profil, idet de seneste tweets udelukkende omhandler spil-opdateringer, hvilket også er illustreret i figuren.



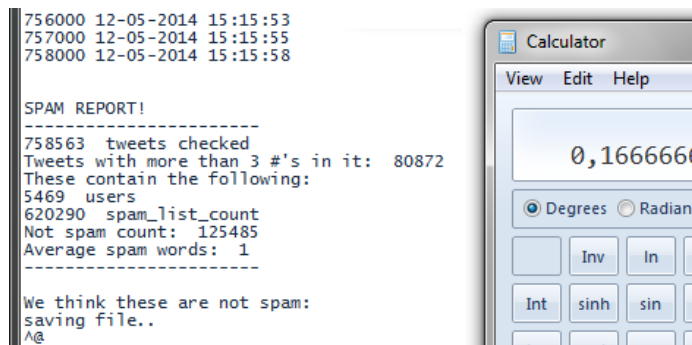
Figur 7.4: Eksempel på en SPAM bruger

Det har ikke været muligt at finde tal på hvor stor en procentdel af alle tweets - generelt set, som kan kategoriseres som spam-relateret. Det har derfor været nødvendigt, selv at udarbejde en definition for tweets som indeholder hvid støj. Følgende er en liste over de elementer, der i dette projekt vil blive kigget på, når der skal frasorteres tweets af hvid støj.

- Ord, som eksisterer i en ordliste med spam-relaterede ord
- Hvis den pågældende tweet indeholder mere end tre hashtags
- Link, follow eller re-tweet ønske

Ordlisten, en stopordliste, er blandt andet bygget op fra observationer i 7.2 samt andre tweets med hvid støj og ses i Bilag D. Ord som *ipad*, *follow* og *iphone* står også på listen. I sig selv er ordene ikke tegn på hvid støj, men hvis der både bliver nævnt *ipad*, *follow*, *coins* og *harvested* stiger sandsynligheden drastisk. Ligeledes kan forekomsten af mange hashtags også være et tegn på hvid støj, ved at "hashjacke" en lang række forskellige brand navne. Et andet eksempel på dette kan være de ponografiske tweets, som nævner en lang række "emner", hvilket så kan ses på det vedlagte link i tweeten. Derudover indeholder tweets med hvid støj et af tre elementer. Et link, som skal klikkes på for mere information, et *follow* ønske eller *re-tweet*. Follow og eller retweet ønske kan derved være klare tegn på hvid støj, men da det ikke er alle der indeholder dette, hjælper det kun til at identificere nogle af dem. De hvid støjs tweets med link benytter ofte også tjenester som *bitly.com*, hvilket skjuler hvor linket fører hen til og derfor ikke kan klassificeres som hvid støj ud fra dette.

For at teste frasortering af tweets med hvid støj bliver der arbejdet på et datasæt af 750.00 tweets. Der bliver talt antal ord fra ordlisten samt antal hashtags.



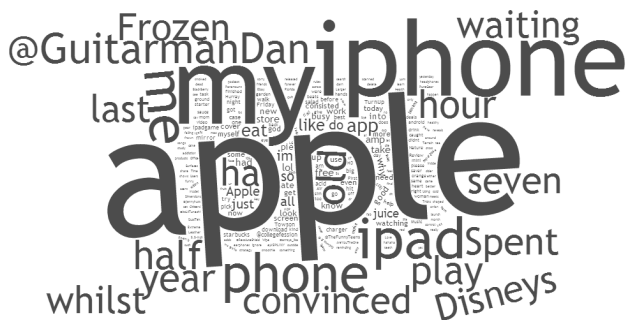
Figur 7.5: Overblik af hvor meget SPAM der er tilstede på brand specifikke søge termer

Resultatet illustreres på figur 7.5 at 620.290 af de 758.563 blev kategoriseret som hvid støj. Lidt over 80.872^2 af disse var med mere end 3 hashtags, hvilket understøtter, at hashtags i sig selv ikke er nok. Derfor er en stopordliste relevant. Ved heuristisk og test blev det klart, at en kombination af antal hashtags og

²80.872 tweets

stop-ord giver et klarere billede af hvilke tweets, der kan kategorisere som hvid støj. På figur 7.5 ses antallet af "rene tweets" hvilket betyder, at kun 16% ikke er hvid støj. Dette lyder som et højt tal. Altså at 84% er ikke ønskede tweets. Det har ikke været muligt at finde statistik på hvor mange procent af tweets er hvid støj, og som First Monday beskriver i en omfattende artikel [29] er dette en svær proces at finde ud af. Da der kun er 140 karakterer at gøre godt med, kan det være svært at fange alle ikke ønskede tweets, og derfor svært at definere hvor mange der er. En ting er dog klart. Det kommer meget an på hvilke søgeord man benytter. I dette tilfælde var de store syndere primært tilknyttet brandet Apple med søgetermerne "ipad" og "iphone"³. Da dette blev klart, var en af overvejelserne at undgå brands som ikke var lige så udsat for *hashjacking*. Dette vil dog på lang sigt frasortere end del fremtidige kunder, og stadig ikke løse problemet med frasortering af hvid støj.

Ved at sammenligne 200 Apple tweets før frasortering, som er illustreret i figur 7.2 er der stor forskel på hvilke ord, der bliver brugt mest. 200 tilfældige apple tweets er vist i illustration 7.6 som viser en stor forskel. I Tabel 1 i Bilag D kan en sammenligning af frekvensfordelingen for figur 7.2 og figur 7.6 ses. Der ikke er så massivt mange ord som *follow*, *retweet* og *rt*, men mere mange forskellige ord, hvilket tyder på almindelig tale kontra korte reklamer.



Figur 7.6: Wordcloud for 200 tweets til brandet Apple uden hvid støj

³tweets som indeholder #movie, #iphone, #android efterfuldt af en lang række pornografiske ord er et typisk eksempel på "hashjacking"

7.4 Evaluering af 3. iteration

I dette kapitel er der ved hjælp af de opstillede hypoteser blevet defineret, hvilke processer der skal gennemløbes, for at filtrere alle unødvendige tweets fra. De unødvendige tweets blev defineret som værende ikke relevante for de udvalgte brands samt hvid støj. Det resterende datasæt vil herved være mere målrettet virksomheder brands.

For at filtrere søgningen på specifikke søgetermer blev der manuelt udvalgt hvilke ord, som matchede de udvalgte brands og derved udvide den relevante søgning.

I forbindelse med denne proces blev der skabt et mindre men mere specifikt datasæt, som var mere målrettet, men stadig indeholdt store mængder hvid støj.

Ved at kigge nærmere på hvordan Twitter håndterer hvid støj blev det klart, at på trods af Twitters egne forholdsregler for at undgå hvid støj, florerer der stadig store mængder hvid støj på Twitter. For at undgå disse, var der behov for at opsætte et hvid støjs filter, for at frasortere ikke relevante tweets.

For at dette kunne ske var det derfor nødvendigt at undersøge, hvilke typer hvid støj der blev observeret i det gemte datasæt, og en definition blev udarbejdet. Da disse var blevet klassificeret, blev der udarbejdet en stopordliste, som i forbindelse med filtrering på antal hashtags udgjorde et modul, som kunne frasortere store mængder af den hvide støj.

Resultatet var, at store mængder data blev frasorteret, da få af de udvalgte brands var udsat for betydelig store mængder ikke ønskede tweets.

Til sidst blev der sammenlignet datasæt inden og efter frasortering af hvid støj. Det blev klart at det nye mindre datasæt indeholdt betydelig flere relevante tweets, i form af brand relateret, hvor unødvendige tweets er frasorteret.

7.5 Validering af forretningsmodel

Det er i denne iteration blevet valideret, at det er muligt at filtrere på specifikke søgetermer samt muligheden for, at kunne sortere hvid støj og andre ikke relevante data fra.

Det er herved påvist, at det er muligt at opfylde grundlag for hhv. et emne filter

og et spam filter beskrevet i forretningsmodellen i afsnit 4.2.1.4. Forretningsmodellen samt de berørte løsninger er illustreret i figur 7.7.

Tagster - Forretningsmodel

<p>PROBLEM</p> <p>At indhente information omkring et specifikt emne</p> <p>At indhente kun relevant data og få frasorteret unødvendig data</p> <p>At indhente holdninger og meninger ud fra et brand</p> <p>EXISTING ALTERNATIVES</p> <p>BrandWatch</p> <p>Overskrift.dk</p> <p>Twitter</p> <p>HootSuite</p> <p>Falcon Social</p>	<p>SOLUTION</p> <p>Et emne-filter, som kan indhente data ud fra et emne</p> <p>Et spam-filter, som kan fjerne al ikke relevant data</p> <p>En metode, således meninger og holdninger kan udtrækkes</p> <p>KEY METRICS</p> <p>Hovedområde: Indhente, analysere/filtrere og præsentere social data</p> <p>Succeskriterie: Opbygge en virksomhed med 10 faste kunder efter første år</p>	<p>UNIQUE VALUE PROPOSITION</p> <p>Gør hvad I gør bedst, og lad os fortælle hvad folk mener om Jer</p> <p>Intet behov for søgeord eller svære forespørgsler, det hele sker automatisk</p> <p>Vi gør social data spiseligt</p> <p>HIGH-LEVEL CONCEPT</p> <p>Social data analyse uden al det unødvendige</p> <p>Social data analyse gjort simpelt</p>	<p>UNFAIR ADVANTAGE</p> <p>CHANNELS</p> <p>Venner og bekendte</p> <p>Mund-til-mund metoden</p> <p>Sociale medier (viral markedsføring) herunder Facebook og Twitter</p> <p>Landing-page</p> <p>AdWords kampagner</p>	<p>CUSTOMER SEGMENTS</p> <p>#1 Mindre webshops</p> <p>#2 Marketingsbureauer</p> <p>#3 Nyhedsbureauer</p> <p>EARLY ADOPTERS</p> <p>Søger indblik i social medie omtale</p> <p>Ønsker en dybere forståelse af hvad folk mener om specifikke emner</p>
<p>COST STRUCTURE</p> <p>Hosting omk. På nuværende tidspunkt gratis lokal server.</p> <p>Udviklingstimer er på nuværende tidspunkt gratis. Forventes at tilbyde partnerskab til eventuelle nye udviklere.</p> <p>AdWords kampagner. Athænger af CTR. Budget sat lavt til 20kr/dag.</p> <p>Fremtidig hosting hos Google Cloud: Mindst kr. 212,39/md. Variabel omk. forbundet med mængde af data indhentet</p>		<p>REVENUE STREAMS</p> <p>#1 "Basis" med 7-dages prøveperiode uden binding, herefter 499kr/md.</p> <p>#2 "Godt klædt på" med 14-dages prøveperiode uden binding, herefter 999kr/md.</p> <p>#3 "Overdrevet" med 30-dages prøveperiode uden binding, herefter 1999kr/md.</p>		

Lean Canvas is adapted from The Business Model Canvas (BusinessModelGeneration.com) and is licensed under the Creative Commons Attribution-Share Alike 3.0 Un-ported License.

Figur 7.7: Forretningsmodel for tredje iteration

4. iteration

Fjerde iteration vil gennemgå det brand-filtrerede data fra kapitel 7, hvor hvid støj blev fjernet. Gennem opstillede hypoteser vil data blive analyseret i form af frekvensfordeling, stemningsanalyse samt rangering og sortering af hhv. brugere og tweets. På denne måde kan man rangere og sortere data i produktet efter relevans.

8.1 Hypoteser

Inden videre analyse af data er det nødvendigt at opstille hypoteser for, hvorvidt det er muligt at indhente relevant information ud fra det indhentede data. Følgende hypoteser er derfor opstillet.

- Det antages at kunne finde relevante og beskrivende ord ud fra et givent brand
- Det antages at kunne vurdere stemningen i en tweet
- Det antages at kunne vurdere en brugers relevans, samt vigtigheden af en tweet

Følgende afsnit vil forsøg at teste de opstillede hypoteser ud fra frekvensfordeling, stemningsanalyse samt rangering og sortering.

8.2 Frekvensfordeling

I kapitel 6 blev det beskrevet, hvordan et datasæt omkring brands blev indhentet. I dette datasæt er det derfor gjort muligt ud fra specifikke forespørgsler at kunne indhente tweets ud fra et givent brand.

Ved at gøre brug af frekvensfordeling¹ er det muligt at se, hvilke ord der fremtræder hyppigst. Ud fra en liste af de mest anvendte ord, vil det være muligt at finde de mest relevante ord for et givent brand. Eksempelvis kan der indhentes data over en vis periode, for herefter at analysere og ud fra frekvensfordeling se, hvilke ord brugerne anvender hyppigst omkring det givne brand.

I forbindelse med dette, vil web-adresser, brugernavne samt hashtags etc. blive fjernet, så det fremadrettet er nemmere at analysere det faktiske indhold i tweet'en. Her kan med fordel anvendes *NLTK's Word Tokenizer*[19], hvor ordene i en tweet vil blive splittet op.

Denne proces er illustreret i figur 8.1, hvor tre tweets har fået fjernet web-adresser, brugernavne samt hashtags, og herefter blevet splittet op pr. ord.

```

RAW: RT @SteveStfler: iâm like an iphone i just lose energy without doing anything
Pretty: rt iâm like an iphone i just lose energy without doing anything
Tokenized: ['rt', 'iâm', 'like', 'an', 'iphone', 'i', 'just', 'lose', 'energy', 'without', 'doing', 'anything']
-----
RAW: Apple scoops up one of Nokia's Lumia camera wizards http://t.co/pyUTGIW6ZC #nokia
Pretty: apple scoops up one of nokia's lumia camera wizards
Tokenized: ['apple', 'scoops', 'up', 'one', 'of', 'nokia', 's', 'lumia', 'camera', 'wizards']
-----
RAW: The Top 5 iPad Apps Being Used In Classrooms Right Now http://t.co/42Wxb0J6H9 via @edudemic #ipaded #tcdsb21c
Pretty: the top 5 ipad apps being used in classrooms right now via
Tokenized: ['the', 'top', '5', 'ipad', 'apps', 'being', 'used', 'in', 'classrooms', 'right', 'now', 'via']
-----

```

Figur 8.1: Eksempler på rå, rene samt tokenized tweets

På denne måde bliver det nemmere at analysere, og ved at fjerne eventuelle stopord, er det muligt at stå tilbage med kun de mest relevante og beskrivende ord. Tidligere er stopord beskrevet som tegn på hvid støj, men i dette tilfælde er der blot tale om fyldord, som ikke bidrager til den videre analyse.

Ved at have fjernet fyldord og overflødige tegn, herunder ;,!, "%', er der kun rå og relevante ord tilbage. Disse ord giver et overblik over de ord brugerne anvender mest i sammenhæng med det givne brand. Et eksempel på de ti mest anvendte ord omkring brands som *Samsung*, *Apple* og *Nokia*, er illustreret i tabel 8.1. Der er her fjernet de respektive brandnavne fra hver liste, da disse ellers vil blive omtalt flest gange.

De tre brands har alle det til fælles at de producerer smartphones. Ved at sætte dem op mod hinanden er det gjort muligt at danne sig et overblik over hvilke

¹Frekvensfordeling fordeling over de ord, som optræder hyppigst.

Samsung	Apple	Nokia
galaxy	iphone	never
phone	ipad	mess
iphone	like	win
note	one	chance
one	im	phone
new	using	amp
phones	phone	lumia
charger	new	support
like	get	bag
usa	store	exclusive

Tabel 8.1: De 10 mest anvendte ord for Samsung, Apple og Nokia

ord de har til fælles, men også deres forskelligheder. Ud fra termerne ses det, at Nokia gør et stort arbejde i at udføre markedsføring på de sociale medier. Her anvendes der konkurrencer, hvor man eksempelvis kan vinde en Lumia-smartphone. Dette ses ud fra ord som *win*, *lumia*, *chance*. Samsung og Apple har ikke behovet for at skulle promovere deres produkter pga deres velkendte brands. Yderligere seks brands og deres mest anvendte ord er illustreret i figur 4 i Bilag D.

For yderligere illustration af de mest anvendte termer er der udarbejdet wordclouds, hvor de hyppigst anvendte ord er størst. I figur 8.2 er der for Apple brandet illustreret de hyppigst anvendte ord, og de resterende brands er vist i Bilag E.



Figur 8.2: Wordcloud for Apple uden ordet *apple* med

Det ses, at ikke kun ord relevante for Apple firmaet indgår i wordcloud'en i figur 8.2. Her optræder ord som "*pie*" og "*juice*", der faktisk er relevant iht. termen *Apple*, men i forbindelse med dette projekt er ordene ikke relevante. Med fordel kunne man foretage en forespørgsel på tweets, som indeholder termet *Apple*, men ikke termer som *pie* og *juice*. Denne metode kan anvendes til at få udarbejdet en bedre liste over relevante ord ud fra et givent brand. Det er hermed bevist,

at det er muligt gennem en række processer at finde relevante og beskrivende ord ud fra et givent brand.

8.3 Stemningsanalyse

En stemningsanalyse er en måde, hvorpå det er muligt ud fra indhentet data, at kunne beskrive humøret for en tweet, og dermed også kunne skabe *mere værdi*. Virksomheder antages at være interesseret i brugernes meninger og holdninger til deres produkter eller services. Ved at anvende stemningsanalyse kan humøret omkring specifikke brands afspejles. Som P. P. Alexander Pak[2] beskriver, er der tre grunde til at foretage stemningsanalyse. Følgende er en liste over de tre grunde.

- Hvilken holdning brugere har til et givent produkt
- Er der tale om en positiv eller negativ holdning om et givent produkt
- Hvad brugere ellers kunne ønske sig af et givent produkt

Ved at se på om brugerne udtrykker sig positivt eller negativt om et brand er det muligt at se, hvad de mener om et givent produkt. Derudover er det i nogle tilfælde også muligt for virksomhederne at finde frem til, hvad kunderne og brugerne ellers kunne ønske sig af produktet. Et eksempel på dette er illustreret i indhentet tweet herunder.

"iphone chargers are so fucking shit!", tweet om Apple.

Brugeren her er tydeligvis frustreret over opladeren til hans iPhone. Denne oplysning kan gavne Apple, da man kan få skabt et overblik og spotte eventuelle problemstillinger. For nyligt blev opladere af ældre dato fra Apple trukket retur[75].

Stemningsanalyse kan udarbejdes på forskellige måder, og kan variere alt efter behov. Sociale medier er kendt for korte tekster, og med Twitter's 140 karakterer er der ikke plads til dybfølte meninger, men primært korte fakta eller holdninger der kommer til kende. Finn Årup Nielsen[52] beskriver, at det kan være svært at bedømme humøret ud fra twitter-lingo.

Der er i dette projekt fremhævet to mulige modeller til stemningsanalyse. Det er forsøgt gennem analyse af en tweet at finde den model, som passer bedst til projektet.

- AFINN[54], stemning beregnet på baggrund af en ordliste med vægtede ord
- NLTKs Naive Bayes[55], stemning beregnet ud fra en trænet model

Følgende afsnit er en gennemgang af hhv. AFINN og Naive Bayes.

8.3.1 AFINN

AFINN[53] er en ordliste, der indeholder 2477 hånd-vægtede ord på en skala fra -5 til 5, hvor -5 er negativt og 5 er positivt. Ved at sammenligne hvert ord i en tweet med ordliste, kan der udregnes en akkumuleret værdi for hele tweet'en. Den akkumulerede værdi fortæller derved om den givne tweet er positiv eller negativ.

Denne metode giver et hurtigt indblik i, hvorvidt en tweet er overvejende positiv eller negativ, men den har dog sine begrænsninger. Idet der kun sammenlignes på ord fra listen, vil det alt andet lige ikke være muligt at finde en værdi for alle ord anvendt i en tweet. Hvis ordet ikke findes i listen, vil ordet få en neutral værdi. Dette starter så endnu et problem, at det ej heller er muligt at spotte ironiske tweets. Et eksempel på disse er illustreret i figur 8.3, hvor fire forskellige tweets er opstillet. Den første tweet er overvejende positiv pga ord som *love*. Den anden tweet vurderes som negativ pga. anvendelsen af to meget negativt-ladete ord. Den tredje tweet er et eksempel på en ironisk tweet. Den vurderes som positiv, men tager man højde for konteksten, skulle tweeten være vurderet enten negativ eller ironisk. Den sidste tweet vurderes som neutral, da et stærk negativt ladet ord, "*overdøver*" de positive ord. Denne tweet burde være angivet som værende positiv.

Som illustreret i figur 8.3 er det ikke muligt med AFINN at kunne vurdere tvetydige tweets. Der er enten tale om negative, neutrale eller positive tweets. Anvendelsen af denne metode giver dog et generelt billede af humøret i en tweet, hvilket er tilstrækkeligt i første version af produktet.

8.3.2 Naive Bayes

NLTK's Naive Bayes[55] er anderledes på mange punkter. Her trænes en model, som herefter kan vurdere stemningen i en tweet. Denne model sørger selv for ikke at skulle vurdere stopord, da modellen automatisk ignorerer disse.

sentiment	text_raw	
3.67	love love love the <u>iphone</u> headphones	
	[3, 3, 3, 0, 0, 0]	
-3.02	<u>iphone</u> chargers are so fucking shit!	
	[0, 0, 0, 0, -4, -4, 0]	
3.74	Absolutely love reaching my data limit on my <u>iphone</u> . So great.	
	Good. Awesome.	
	[0, 3, 1, 0, 0, 0, 0, 0, 0, 0, 3, 3, 4, 0]	
0.0	Fuck me! I love my new <u>iphone</u> so much! makes me feel alive	
	[-4, 0, 0, 3, 0, 0, 0, 0, 0, 0, 0, 0, 1]	

Figur 8.3: Sentiment Analyse på fire Eurovision sætninger

I AFINN gøres modellen mere præcis ved at tilføje flere ord til listen, hvor man med Naive Bayes skal træne modellen. Processen for dette gennemgår Stephen Whitworth's i artiklen[82]. Her gør man brug af hhv. ti positive og negative stemte sætninger, der anvendes som træningsdata. Disse er illustreret i figur 8.4. Et større træningssæt vil være at foretrække, som han også nævner i artiklen.

Positive sætninger	Negative sætninger
I love their new album	This band is awful
I want to go and see them live	This is terrible
I like this band	They make me feel bad
I love the lead singer he is so dreamy	This band are not that great
I adore them they are so great	This is absolutely atrocious
They are brilliant	I dislike them
They are the best	This band are the worst
I love this band so much	What the hell is this crap
I think this band are great	This band should die
This band for president	I hate this band

Figur 8.4: Træningssæt bestående af 10 positive og 10 negative sætninger. Kilde: [82]

Ved at benytte træningssættet og anvende det indhentede datasæt med 3,4 mio. tweets beskrevet i afsnit 6.2, var det muligt at træne modellen. Gennem en længere iterativ proces blev der trænet 544 positive, 620 negative og 488 neutrale tweets. Dette vurderes jf. Whitworth[82] til at være et tilstrækkeligt træningssæt til videre brug. Ulempen ved denne model er, at modellen ikke er klogere end det data man træner den med, og at træne modellen tager lang tid. Til gengæld vil modellen give et mere præcist resultat end set hos AFINN i afsnit 8.3.1. En mulighed for at gøre træningssættet bedre, ville være ved hjælp af *crowd sourcing*. Her vil man gennem produktet kunne få feedback på stemningen

af hver tweet. Dermed kan brugerne selv bidrage til et bedre system.

Ved at gøre brug af AFINN gives en simpel og generel vurdering af stemningen i en tweet, hvor Naive Bayes vil give et mere korrekt billede. Over tid, vil Naive Bayes give et bedre resultat, men på nuværende tidspunkt findes AFINN tilstrækkelig for projektet.

8.4 Rangering og Sortering

Det er i dette projekt relevant at kunne rangere og sortere hhv. brugere og tweets, således at kun relevante bliver præsenteret i produktet. For at kunne rangere både brugere og tweets for herefter at sortere på dem efter relevans, er det nødvendigt at få opstillet en definition af disse. I de følgende afsnit vil de blive defineret samt illustreret med eksempler.

8.4.1 En relevant bruger

Først og fremmest er det essentielt at en relevant bruger ikke genererer *hvid støj*, som tidligere beskrevet i afsnit 7.3. For videre at vurdere hvad der gør en bruger relevant, er følgende parametre fundet relevante.

- Antallet af følgere
- Klout score^[44]
- Antal tweets, samt andelen af retweets og favorites

Antallet af følgere, beskriver antallet af personer, som lytter til brugeren. Det er her antaget, at antallet af følgere er proportionalt med hvor mange personer, som lytter til de tweets han skriver. Hvis en bruger har mange følgere, så må det alt andet lige betyde, at brugeren er betydningsfuld for følgerne.

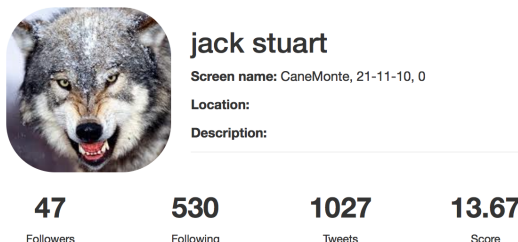
Klout score, beskriver brugerens indflydelse på en skala fra 1-100 jf. Klout. Klout beskriver indflydelse, som effekten af en opdatering, der får følgerne til at reagere. En effekt kunne eksempelvis være, at brugerens følgere delte brugerens tweet. Har brugeren stor indflydelse blandt sine følgere, så vil hans klout score være i den høje ende af skalaen.

Antal tweets, samt andelen af retweets og favorites, beskriver vigtigheden for hver tweet, brugeren skriver. Sammenlagt med antallet af tweets brugeren har skrevet, vil det være muligt at finde et gennemsnit for retweets samt favorites, som direkte beskriver effekten eller indflydelsen på brugerens tweets.

Et eksempel på hhv. en relevant og knapt så relevant bruger er illustreret i figur 8.5 og figur 8.6.



Figur 8.5: En relevant bruger. Kilde: <http://beta.tagster.dk/user/barackobama/>



Figur 8.6: En knapt så relevant bruger. Kilde: <http://beta.tagster.dk/user/canemonte/>

Det ses, at brugeren illustreret i figur 8.5 har et højt antal følgere og ud fra klout scoren må det antages, at brugeren har stor indflydelse på følgerne. Lige modsat er det illustreret i figur 8.6, hvor brugeren har et relativt lavt antal følgere, hvilket også afspejler sig i brugerens klout score.

I forbindelse med dette projekt er valget af klout fundet, blandt andre alternativer[23], som en løsning. Fremfor at skulle indhente *antallet af følgere, antallet af tweets*, samt beregne andelen af *retweets og favorites*, så kan en enkelt forespørgsel til Klout give et overslag på brugerens relevans. Som udgangspunkt i dette projekt vil en klout score på over 50 være med til at kategorisere brugeren som værende relevant. Dette kan senere hen justeres alt efter behov. Et eksempel på en gennemsnitlig aktiv bruger med en klout score på over 50 er illustreret i figur

8.7.



Figur 8.7: En gennemsnitlig relevant bruger. Kilde: <http://beta.tagster.dk/user/UlrikUhreBrink/>

8.4.2 En vigtig tweet

Hvert sekund bliver en ny tweet skrevet, og som beskrevet i afsnit 2.4, så bliver der hver dag skrevet flere millioner tweets. Det er derfor nødvendigt at definere vigtigheden af en tweet, således at de mest relevante bliver rangeret højest. Følgende parametre er fundet relevante, når vigtigheden af en tweet skal findes.

- Hvem der har skrevet den
- Positiv eller negativ stemt
- Antallet af retweets/favorites

Hvem der har skrevet den, er en vigtig faktor for en tweet. Det er beskrevet i afsnit 8.4.1, hvordan indflydelsen af en bruger har effekt på spredningen af budskabet i en tweet. Hvis en bruger har stor indflydelse blandt hans følgere, så vil sandsynligheden for, at en tweet bliver spredt hurtigt være høj.

Om en tweet er *positiv eller negativ stemt*, kan have stor betydning, især hvis denne bliver delt et højt antal gange eller ligefrem går viralt. Hvorvidt en tweet er positiv eller negativ stemt beskriver humøret for brugeren, da den blev skrevet. Hvis et produkt er blevet omtalt negativt og er kommet ud til en stor mængde brugere, vil denne skulle angives som vigtig, således der kunne tages hånd om en evt. problematik. Det antages, at især positive tweets kan være tegn på ambassadører, som promoverer et produkt, uden at få betaling. For at denne tweet skal rangeres som værende vigtig, er det nødvendigt, at brugeren har en høj nok indflydelse blandt følgerne, hvilket er beskrevet i afsnit 8.4.1.

Antallet af retweets og favorites, kan på mange måder beskrive, hvor populær en tweet er. Hvis en tweet er blevet delt mange gange, så har den alt andet lige et højt antal retweets og favorites, hvilket direkte afspejler sig i populariteten blandt brugerne på Twitter.

Et eksempel på en hhv. vigtig og knap så vigtig tweet er illustreret i figur 8.8 og figur 8.9. Her anvendes *#LM24* i begge tweets, hvor den vigtige har både *retweets* og *favorites*, og den mindre vigtige ingen har.



Figur 8.8: En vigtig tweet



Figur 8.9: En knap så vigtig tweet

I forbindelse med at finde en vigtig tweet, kan følgende citat anvendes.

„*Hvis et træ falder i skoven. Har det så en lyd?(oversat)*“ [4], Berkeley (1710).

Citatet bliver anvendt i mange henseender, og i dette projekt vil følgende udgave sammen med de opstillede parametre, beskrive en vigtig tweet.

Hvis en bruger skriver en tweet, og ingen lytter eller følgere har. Har han så en stemme?

Det korte svar er *nej*. Hvis brugeren deler sit budskab gennem sine tweets men ingen følgere har, så er sandsynligheden for at nogen hører hans stemme minimal. Brugerens tweet kan derfor ikke antages at være vigtig i forhold til et givent brand. Hvis brugeren havde haft flere følgere, så ville flere personer alt andet lige lytte til hans budskab, og givet vis havde budskabet spredt sig. Har brugeren omtalt et produkt eller en service enten positivt eller negativt, vil denne ud fra de opstillede parametre, blive antaget som værende vigtig. Dette sker bl.a. på baggrund af brugerens indflydelse på hans følgere.

Ved at gøre brug af *call to action*, kan særligt vigtige tweets fremhæves, hvor virksomheden bør reagere. Er der tale om særligt positive tweets, kan man med fordel gøre brugeren til ambassadør for produktet, hvilket afhænger af brugerens relevans. Er der derimod tale om særligt negative tweets, kan virksomheden gå ind i samtalen og få løst et eventuelt problem. Er der tale om mange negative tweets over en kort periode, kan der være tale om et generelt problem. Her kan virksomheden reagere hurtigt, og inden et faktisk problem opstår, så har virksomheden løst det.

8.5 Evaluering af 4. iteration

I dette kapitel blev der opstillet tre hypoteser for at kunne finde relevante og beskrivende ord omkring et brand, vurdere stemningen i en tweet, samt vurdere hhv. en brugers relevans og vigtigheden af en tweet.

Gennem frekvensfordeling blev det gjort muligt, at finde relevante og beskrivende ord omkring et brand, baseret på hyppigst anvendte ord brugerne anvender. I forbindelse med dette, blev *NLTK's Word Tokenizer* anvendt til at splitte en tweet op, således at web-adresser, brugernavne og hashtags samt stopord kunne undlades. Ud fra ni opstillede brands, blev relevante og beskrivende ord fundet.

En stemningsanalyse kan bestemme en brugers holdning til et produkt, om det forholder sig positivt eller negativt. Der blev testet to modeller til at foretage stemningsanalyse - AFINN og Naive Bayes. Her blev AFINN fundet som den mest relevante hvad angår dette projekt, selvom Naive Bayes over tid vil give et mere præcist resultat.

For at kunne rangere og sortere på hhv. brugere og tweets, blev der opstillet en definition af en relevant bruger samt en vigtig tweet. Det blev fundet, at en relevant bruger er defineret ud fra *antallet af følgere, klout score og antal tweets, samt andele af retweet og favorites*. En vigtig tweet blev defineret ud fra *brugeren, som havde skrevet den, om den var positiv eller negativ stemt og ud fra antallet af retweets og favorites*. Med udgangspunkt i en vigtig tweet og en relevant bruger, blev det gjort muligt at fremhæve disse. Dette blev illustreret gennem *call to action*, der opfordrer virksomheder til at reagere der på.

8.6 Validering af forretningsmodel

Det er i denne iteration blevet valideret, at det er muligt ud fra brand specifik data at vurdere stemningen og yderligere opdele data efter relevans.

Det er herved påvist at det er muligt, at udarbejde en metode hvorledes meninger og holdninger kan udtrækkes, beskrevet i forretningsmodellen i afsnit 4.2.1.4. Forretningsmodellen samt den berørte løsning er illustreret i figur 8.10.

Tagster - Forretningsmodel				
PROBLEM At indhente information omkring et specifikt emne At indhente kun relevant data og få trasoreret unødvendig data At indhente holdninger og meninger ud fra et brand EXISTING ALTERNATIVES BrandWatch Overskrift.dk Twitter HootSuite Falcon Social	SOLUTION Et emne-filter, som kan indhente data ud fra et emne Et spam-filter, som kan fjerne al ikke relevant data En metode, således meninger og holdninger kan udtrækkes	UNIQUE VALUE PROPOSITION Gør hvad I gør bedst, og lad os fortælle hvad folk mener om Jer Intel behov for søgeord eller svære forespørgsler, det hele sker automatisk Vi gør social data spiseligt	UNFAIR ADVANTAGE	CUSTOMER SEGMENTS #1 Mindre webshops #2 Marketingsbureauer #3 Nyhedsbureauer EARLY ADOPTERS Søger indblik i social medie omtale Ønsker en dybere forståelse af hvad folk mener om specifikke emner
	KEY METRICS Hovedområde: Indhente, analysere/filtrere og præsentere social data Succeskriterie: Opbygge en virksomhed med 10 faste kunder efter første år	HIGH-LEVEL CONCEPT Social data analyse uden al det unødvendige Social data analyse gjort simpelt	CHANNELS Venner og bekendte Mund-til-mund metoden Sociale medier (viral markedsføring) herunder Facebook og Twitter Landing-page AdWords kampagner	
COST STRUCTURE Hosting omk. På nuværende tidspunkt gratis lokal server. Udviklingstimer er på nuværende tidspunkt gratis. Forventes at tilbyde partnerskab til eventuelle nye udviklere. AdWords kampagner. Afhænger af CTR. Budget sat lavt til 20kr/dag. Fremtidig hosting hos Google Cloud: Mindst kr. 212,39/md. Variabel omk. forbundet med mængde af data indhentet		REVENUE STREAMS #1 "Basis" med 7-dages prøveperiode uden binding, herefter 499kr/md. #2 "Godt klædt på" med 14-dages prøveperiode uden binding, herefter 999kr/md. #3 "Overdrevet" med 30-dages prøveperiode uden binding, herefter 1999kr/md.		

Lean Canvas is adapted from The Business Model Canvas (BusinessModelGeneration.com) and is licensed under the Creative Commons Attribution-Share Alike 3.0 Un-ported License.

Figur 8.10: Forretningsmodel for fjerde iteration

5. iteration

Femte iteration vil illustrere hvordan der er udarbejdet to MVP'er. Den ene skal danne grundlag for præsentationen af det relevante data indhentet i kapitel 6, filtreret og sorteret i kapitel 7. Den anden skal forsøge at promovere produktet, og indhente oplysninger om hvorvidt, der faktisk er efterspørgsel på denne type produkt på markedet.

9.1 Hypoteser

Inden udarbejdelsen af de to MVP'er er det nødvendigt at opstille hypoteser for at kunne teste, hvorvidt det er en succes eller ej. På baggrund af disse hypoteser er det muligt at se, hvorvidt der er belæg for videre udvikling eller ej.

Hypoteser

- Det er antaget, at der på markedet for sociale medier et behov for dette produkt
- Det er antaget, at mængden af funktioner er tilstrækkelig til første udgave af produktet, som leveres til kunderne og tages betaling for

9.2 To typer af MVP

I dette projekt er der gennemgående arbejdet efter lean startup principperne, som beskrevet i kapitel 4. Fælles for dette er, at man ikke udvikler et færdigt

produkt før man er sikker på, at der faktisk er efterspørgsel på produktet. I dette projekt er der derfor udarbejdet to forskellige MVP'er, som hver især indeholder forskellige core-features, som man kan gå til markedet med. De to typer er følgende.

- En landing-page
- Web-applikation (Udvidet Wizard of Oz)

Anvendelsen af en landing-page gør det muligt at få udarbejdet et sted (en hjemmeside), hvor besøgende vil kunne finde alle informationer omkring projektet - hvad det handler om, hvorfor man har behov for netop dette, og ikke mindst, hvordan man bliver tilmeldt. Landing-pagen kan anvendes til at indhente informationer omkring hvor efterspurgt produktet er, for på den måde at kunne validere værdigrundlaget for projektet, og om hvorvidt ydelserne er prissat korrekt. Denne MVP er nærmere beskrevet i afsnit 9.3.

En anden type der i dette projektet er valgt, er en udvidet "Wizard of Oz"[8] web-applikation. Der er taget udgangspunkt i en "Wizard of Oz", som ligner og opfører sig som et færdigt produkt, men er styret af enten automatiserede processer eller personer. Det er stik imod lean og anvendelsen af MVP, at foretage nogen form for kodning, da det er anset som værende for dyrt i et lean startup, som Ramli John nævner i artiklen[42]. I web-applikationen er der blevet udviklet en platform, hvor kun de mest nødvendige funktioner er implementeret - derfor den udvidede form af "Wizard of Oz", idet web-applikationen opfører sig som et færdigt produkt, og indeholder kun de mest nødvendige funktioner, som man kan gå til markedet med. Denne MVP er nærmere beskrevet i afsnit 9.4.

9.3 Landing-page

Helt basalt er en landing-page en webside, hvor mulige interessede besøgende skal overtales, gennem promovring af produktet, til at blive kunder. I forbindelse med dette projekt er denne landing-page anvendt til at afsøge markedet, for på den måde at se, hvorvidt produktet er efterspurgt eller ej. Gennem en *call-to-action*, herefter omtalt som *cta*, er det muligt at indhente information om, hvor mange brugere der trykker på de forskellige "knapper" på siden. Der er bevidst fravalgt generelle tekster såsom *sign-up* eller *create free account*, men i stedet anvendt et mere almindeligt sprog, og forhåbentligt kunne bidrage til at konvertere bedre. Informationerne indhentet fra aktiviteterne på knapperne

vil danne grundlag for hvor efterspurgt produktet er, og hvorvidt man bør gå videre med produktudviklingen.

9.3.1 Landing-page cases

Der findes forskellige typer af landing-pages, og i forbindelse med dette projekt er der taget udgangspunkt i den minimalistiske type. Særligt i denne kategori kan Dropbox[21] og Mailbox[47] nævnes som eksempler på, hvordan en landing-page promoverer et givent produkt eller service og konverterer "besøgende" til kunder.

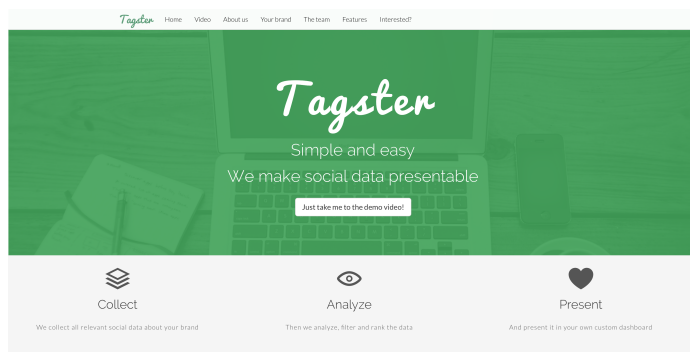
Fælles for **Dropbox** og **Mailbox** er, at de begge er yderst minimalistiske og har et givent mål - at få de besøgende til at oprette sig som kunder hos dem. Dette sker gennem et primært og sekundært cta, og i disse to cases et hhv. "sign-up" knapper, hvor "call"er, at kunden opretter sig, og hvor "action"er, at man klikker på knappen. I disse to cases er der særligt anvendt både primære og sekundære *call-to-actions*, som er understøttet af en visualisering af deres produkt gennem enten illustrationer eller video. Eksempler på disse cases er illustreret i Bilag F.

9.3.2 Gennemgang af landing-page

Landing-pagen er delt op i forskellige blokke, som hver især har til formål at komme med forskellige budskaber. Hver blok er brudt af en såkaldt divider, hvis formål er at opretholde interessen, når der skiftes mellem blokke og deres forskellige typer af indhold. Disse dividers er nærmere beskrevet i afsnit 9.3.2.1.

De inkluderede blokke på siden er følgende, og er nærmere beskrevet herefter.

- Header/appetitvækker
- Video præsentation af hvad man får, samt hvem og hvad Tagster er
- En opfordring til hvorfor man bør *lytte* til sociale medier
- Personerne bag Tagster (relations-faktor)
- Features uddybet og illustreret
- Prissætning
- Kontaktoplysninger



Figur 9.1: Tagster.dk - Header/appetitvækker

Headeren er det første de besøgende bliver mødt med. Denne indeholder logo for produktet, en teaser-tekst for, hvad der tilbydes samt en cta i form som en knap, som henviser til videoen. Da headeren er det første man møder, er det essentielt at skabe blikfang og lysten til at vide mere om produktet. Yderligere i denne blok er illustreret de tre steps processen for produktet gennemgår - fra indsamling af data til præsentation på et personaliseret dashboard. Headeren samt cta-knappen er illustreret i figur 9.1.

Videopræsentationen indeholder en gennemgang af produktet. Her vises de forskellige elementer og funktioner produktet indeholder. Der er endvidere anvendt baggrundsmusik for at lette stemningen og på den måde gøre videoen mere indbydende. En gennemgang fra EyeViewDigital[18] viser, at anvendelse af video på en landing-page kan øge konverteringen med op til 80%, og som Unbounce[77] også nævner det, så bidrager en video også med følgende.

- At holde de besøgende på siden i længere tid, så man kan huske og genkende produktet efterfølgende
- At det er nemmere at se videoen fremfor at bruge tid på at læse tekst og indhold på siden.

Denne blok er illustreret i figur 17 i Bilag G.

Opfordringen ligger i den opstillede teaser, som understøttes af en beskrivende tekst om, hvorfor man bør lytte til de sociale medier. En cta-knap er placeret nederst i blokken, som opfordrer de besøgende til at starte med at lytte allerede idag. Denne blok er illustreret i figur 18 i Bilag G.

Personerne bag er udelukkende med for at skabe en personlig relation til produktet. Denne kunne eventuelt udskiftes med en præsentation af eksisterende kunder, som eventuelle nye kunder bedre vil kunne relatere sig til.

Dette bør også understøttes af en eventuel personlig citering af kunderne, idet det antages, at kunder lytter til andre kunder. Denne blok er illustreret i figur 19 i Bilag G.

Features er først illustreret i demo-videoen, men i denne blok yderligere beskrevet. Systemet er delt op i tre - login, dashboard samt brugerprofil. Ved tryk på en af disse dele åbnes en såkaldt modal¹, hvori de forskellige elementer og features er yderligere beskrevet. Denne blok er illustreret i figur 20 og 21 i Bilag G.

Prissætningen kan opstilles på forskellige måder og det handler i bund og grund om at lede de besøgende hen på netop det produkt, der ønskes solgt. Jf. prisstrukturen i afsnit 4.2.1.6 er der opstillet tre forskellige produkter samt et special-tilpasset produkt, hvoraf et af produkterne ønskes solgt oftere end de andre. For at skabe blikfang og lede de besøgende hen på det ønskede produkt, er der valgt en bestemt placering og farve som indikation. Produktet er placeret mellem to andre produkter - et basis og et mere overdrevet produkt. De er opstillet på en måde, så man ikke finder basisproduktet tilstrækkeligt, og at man finder det mere overdrevne produkt for dyrt - det er ihvertfald tanken, og vha. cta-knapper under hvert produkt, er det muligt at indhente data på, hvilke produkter de besøgende vælger. Hvis ingen af de tre produkter findes tilstrækkelige, så er der yderligere opsat en cta-knap under de andre produkter, hvis ingen af de tilbudte produkter er ønsket. Denne blok er illustreret i figur 22 i Bilag G.

Kontaktoplysninger findes i den sidste blok på siden, og indeholder generelle kontaktoplysninger. Desuden er det muligt at skrive sig op til nyhedsbrevet, hvis nyheder ønskes før alle andre. Dette er igen blot en cta-knap, hvor der indhentes informationer omkring hvor mange, som faktisk ønsker at vide mere og gerne før alle andre. Denne blok er illustreret i figur 23 i Bilag G.

9.3.2.1 Dividers

Idet en landing-page er opbygget af en lang side med forskellige blokke, er der anvendt dividers for at bryde blokkene imellem. Derudover skal disse også bidrage til at opretholde interessen for de besøgende, når de forskellige blokke læses. Hver divider's indhold relaterer sig til den næste bloks indhold. En divider indeholder en teaser tekst, som også er anvendt i headeren, samt en underbyggende tekst. Hver divider illustreret som et udvalgt billede, som passer til indholdet. Eksempelvis er teaser teksten i den første divider "social data har ingen grænser"(oversat), og dette er illustreret med et billede af det åbne hav. Den

¹Et pop-up vindue - <http://getbootstrap.com/javascript/#modals>

underbyggende tekst er "*social data er som den vilde vesten, udforsket men fyldt med muligheder*" (oversat), dette er herefter underbygget med, at "*vi filtrerer, og udtrækker det relevante data, så du kan udforske det*" (oversat). Flere af disse dividers er illustreret i Bilag G.

9.3.3 Indhentning af data

Hvis det ikke er målbart, så er det umuligt at se, om produktet er en succes, og uden disse data er det umuligt at vide, hvorvidt man bør gå til markedet med ideen, eller allerede nu opgive den.

Der er gennemgående anvendt kortlinks vha. tjenesten bitly[5] i forbindelse med at indhente antal klicks på de forskellige cta-knapper på siden, men yderligere anvendt Google Analytics[34] til at måle, hvor mange der besøger siden og hvordan de er blevet henvist til siden eksempelvis via den oprettede Twitter-profil².

Fra dag 1 og igennem projektførelsen er der løbende delt billeder og kommunikeret budskabet med produktet ud til følgerne af Twitter-profilen. Over denne periode er der blevet indhentet statistik fra hhv. knapper og antal besøgende, som da vil bidrage til grundlaget for videreførelsen af projektet. Resultaterne for Twitter-profilen, samt statistik fra landing-pagen er nærmere beskrevet i afsnit 9.3.3.1.

9.3.3.1 Resultater

Twitter-profilen blev i medio februar oprettet, og de første tweets omhandlede "lækkede" billeder af produktet, og gennem andre netværk³, blev disse tweets delt, i håbet om at nogle ville finde det interessant. April måned var den måned, hvor budskabet kom ud. Nye følgere kom til, hvilket er illustreret i figur 24 i Bilag H, og blandt disse var to personer⁴, som står bag 140tegn.dk. De følger idag udviklingen af produktet på sidelinjen og kommer både med inputs, samt flittigt deler budskabet om produktet. Dette er illustreret i figur 9.2, hvor en bruger⁵ er på udkig efter et produkt, som kan give kvalitativ analyse af hvad der bliver sagt på twitter. I svaret til brugeren, blev produktet nævnt⁶ som værende en mulig løsning til denne problemstilling.

²Tagsterdk - <https://twitter.com/tagsterdk>

³Gennem @UlrikUhreBrink og @thekc114

⁴@lassekjaertw og @lassefrom

⁵@bavnhoej

⁶Nævnt som reference til twitter-profilen



Figur 9.2: @Tagsterdk nævnt som social medie virksomhed

Der blev med det samme aktivt gået ind i samtalen og brugeren, som efterspurgte produktet følger idag stadig twitter-profilen. Twitter-profilen har i skrivende stund over 60⁷ følgere, hvor nogle er mere aktive end andre.

Twitter-profilen har blandt andet stået som kommunikationskanal til at skabe trafik til landing-pagen, som i periode er blevet flittigt besøgt. Næsten 400⁸ besøgende har i perioden besøgt tagster.dk og i tabel 9.1 er en oversigt over, hvilke knapper de besøgende har trykket på.

Placering	Call-to-action	Antal klik
YOUR BRAND, THEIR VOICE.	Start listening today!	18
WE CREATE COOL STUFF	Get your own dashboard today!	17
WE ARE DIRT CHEAP	Get a basic account!	1
WE ARE DIRT CHEAP	Let's suit up!	4
WE ARE DIRT CHEAP	Get more than you need!	3
WE ARE DIRT CHEAP	Get a custom dashboard!	4
THANKS FOR VISITING US	Get news from us!	4

Tabel 9.1: Antal klik ud fra *call-to-action*

De to første *call-to-actions* knapper har som udgangspunkt været dem, som har konverteret mest, hvis man skal bedømme ud fra antal klik. Disse er begge

⁷@Tagsterdk følges af 61 personer pr. 14-06-2014

⁸390 besøgende d. 14-04-2014

illustreret i hhv. figur 26 og 27 i Bilag H. Da disse har en langt større *hit-ratio*⁹ end de andre, kan beskrive følgende to ting. Den ene er, at ud fra den beskrevne tekst om at virksomheder bør lytte til deres kunder på de sociale medier, at de godt kan se fordelene i det, og antallet af klik kan beskrive, at de er interesseret i at lytte til dem. Den anden er, at de finder produktet interessant, og gerne vil have deres eget, så de kan begynde at lytte til deres sociale omtale.

At der ikke er så mange hits på de andre *call-to-actions* kan betyde, at placering for disse er uhensigtsmæssig. En af de besøgende nævner, at det efter opbygget interesse kan være svært, hurtigt at kunne skrive sig op, hvilket er illustreret i figur 25 i Bilag H. Det er essentielt at kunne få de besøgende til at skrive sig op, så snart de har fundet det interessant., uden at de skal bruge tid og kæfter på, at lede efter en *sign-up* knap. At udarbejde et ny landing-page layout og herefter foretage A/B-splittestning[84] vil kunne beskrive, hvorvidt det faktisk er tilfældet. Her vil indhentet data fra begge udgave blive indhentet og stillet op mod hinanden. Hermed vil det være muligt at kunne se hvilken af udgaverne, som konverterer bedst.

Den udarbejdede video beskrevet i afsnit 9.3.2, har vist sig at være særligt attraktiv, og påstanden om at den konverterer op mod 80% kan derfor godt antages at være korrekt. Videoen er i perioden blevet set af næsten 80¹⁰ besøgende. At næsten 1/4 af alle besøgende har afspillet videoen illustrerer deres interesse i produktet og deres nysgerrighed om at ville vide mere om produktet. Resultaterne for videoen er illustreret i Bilag H.

9.4 Web-applikation

Web-applikationen har til formål at præsentere relevante tweets samt andet analyseret data, som er indsamlet, filtreret og sorteret ud fra brand termer. Derudover er en af web-applikationens væsentligste opgaver at skabe overblik over det indsamlede data på en måde, så kun nødvendig og relevant information bliver præsenteret for brugeren. Da der er tale om et udvidet MVP, så er det kun de mest nødvendige funktioner, som er kodet og resten er blot illustreret med fiktive data, for stadig at kunne illustrere et "Wizard of Oz"-produkt. Web-applikationen er illustreret i Bilag I.

Til applikationen er der anvendt Django Web Framework[30] til den bagvedliggende logik, samt en forbindelse til en MySQL-database, hvor al den rå data om

⁹Blevet klikket på flere gange

¹⁰75 afspilninger pr. d. 14-04-2014

samtlige brands ligger gemt. Disse data vil blive præsenteret gennem en HTML5-side, som er designet og bygget vha. Bootstrap[6], så data kan præsenteres på bred vifte af platforme, herunder desktop, tablets og mobile enheder.

Applikationen indeholder forskellige informationer og i denne iteration vil informationerne på siden inkludere analyseret data omkring et brand, en given twitter brugerprofil og finde seneste tweets ud fra et valgt term. De forskellige typer af data er repræsenteret og overvejelserne omkring hvert element er beskrevet yderligere i afsnit 9.4.1.

9.4.1 Gennemgang af web-applikation

Web-applikationen er opbygget af tre forskellige applikationer, som hver især præsenterer forskellige typer af data, som samlet set danner rammen for den samlede web-applikation. De forskellige applikationer er listet herunder, og er yderligere beskrevet i afsnit 9.4.1.1, 9.4.1.2 og 9.4.1.3.

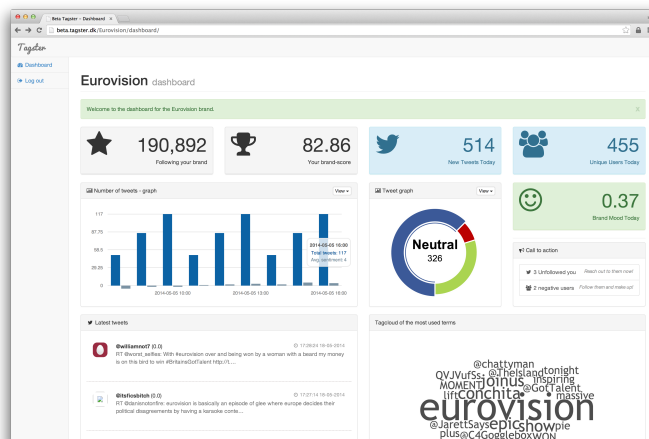
- Dashboard
- Brugerprofil
- Udforsk

9.4.1.1 Dashboard

Dashboardet, som er illustreret i figur 9.3, giver overblik og indsigt i et brands sociale omtale på de sociale medier. Det indhentede og analyserede data er præsenteret og visualiseret gennem forskellige moduler, så kun relevant information er vist.

Dashboardet er overvejende delt op i tre områder, alt efter hvilken type data det drejer sig om. De tre områder er listet herunder og opdelingen farveinddelt og illustreret i figur 31 i Bilag I.

- Generel brand og daglig tweet information
- Brand "humøret"
- Seneste tweets og de mest anvendte termer



Figur 9.3: Tagster.dk - Eurovision Dashboard

Generelle oplysninger indeholder oplysninger om hvor mange personer, der følger det respektive brand på twitter, hvor høj en score¹¹ deres profil har. Derudover indeholder dette område også information om hvor mange nye tweets om det respektive brand, der er blevet skrevet på twitter den pågældende dag, samt antallet af unikke brugere, som har omtalt brandet.

Brand humøret bliver visualiseret på forskellige måder. Den første graf illustrerer antallet af tweets de seneste par timer samt hvordan humøret har været disse timer. Dernæst et cirkel-diagram over samtlige tweets for den pågældende dag inddelt efter humør - positivt, negativt og neutralt. Baseret på samtlige tweets den pågældende dag, er der beregnet det gennemsnitlige humør, og illustreret hhv. som en positiv, negativ eller neutral smiley. Hvis brugere enten stopper med at følge brandet eller en bruger med mange følgere skriver negativt om brandet, så vil dette blive præsenteret i "call-to-action"boksen.

De seneste tweets og de mest anvendte termer er illustreret som en tidslinje og en tagcloud, hvor de mest hyppige termer er størst. Tidslinjen giver et overblik over de seneste tweets, hvad de har skrevet og hvem der har skrevet dem. Hver tweet er yderligere blevet farvelagt, alt efter om der er tale om en positiv, negativ eller neutral tweet. Ved et klik på brugerens profil-billede, åbnes brugerprofilen og denne er yderligere beskrevet i afsnit 9.4.1.2. Ved brugerens navn er yderligere præsenteret brugerens score¹², for på den måde at indikere, hvorvidt der er tale om en indflydel-

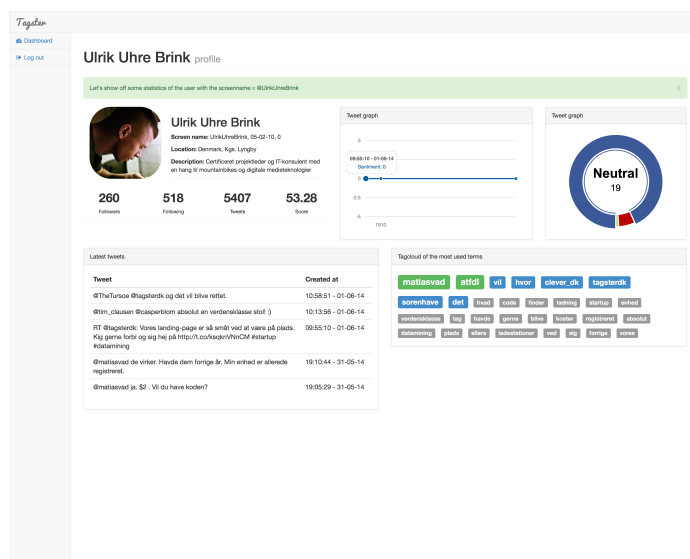
¹¹Et brands score bliver hentet via Klout.com

¹²Hentet vha. Klout API

sesrig bruger eller ej. Hvis brugeren er indflydelsesrig og skriver negativt om brandet, så vil denne tweet være illustreret i "call-to-action"området, som beskrevet tidligere.

9.4.1.2 Brugerprofil

Som beskrevet i afsnit 9.4.1.1, så anvendes denne applikation til at få skabt et overblik over en given bruger. Brugerprofilen har samme opbygning som dashboardet, og indeholder derfor både generelle oplysninger om brugeren, men også analyseret data ud fra de seneste tweets. Brugerprofilen er illustreret i figur 9.4.



Figur 9.4: Tagster.dk - Brugerprofil - UlrikUhreBrink

Generelle oplysninger hentes fra brugerens Twitter-profil, hvorfor antallet af følgere, hvor mange personer brugeren følger, det totale antal tweets brugeren har skrevet, samt brugerens score er illustreret. Disse data skal forsøge at beskrive, hvorvidt brugeren er en person andre lytter til eller ej. Dette baseres på antallet af følgere, hvis brugeren har mange følgere, vil der angiveligt være mange personer, som lytter til brugeren, og derfor er der tale om en indflydelsesrig bruger.

Humøret i de seneste tweets er illustreret på samme to måder, som set i dashboardet. En graf viser humøret i de seneste tweets, og denne graf

giver et godt overblik over, hvornår brugeren fremtræder enten positiv eller negativ, for på den måde at finde en eventuel årsag til, at brugeren fremtræder negativ. Inddelingen af positive, negative samt neutrale stemte tweets er illustreret i et cirkel-diagram, hvor det er muligt at få vist eksempelvis positive tweets.

De seneste tweets og de mest anvendte termer er illustreret som en tidslinje og en tagcloud, hvor de hyppigst anvendte termer er illustreret både efter størrelse og farve. De mest anvendte termer er illustreret med farven grøn, de mindre anvendte i farven blå og termer kun nævnt en enkelt gang er illustreret med farven grå. Ved et enkelt tryk på en given term i tagcloud'en, er det muligt at udforske, hvad andre brugere på Twitter skriver om det givne term. Denne funktion er nærmere beskrevet i afsnit 9.4.1.3.

9.4.1.3 Udforsk

Som applikationens navn også lægger op til, så tillader denne at udforske hashtags og andre termer direkte fra Twitter. Der er blevet udarbejdet en metode så der via Twitter kan hentes tweets ud fra følgende parametre, og dermed udforske, hvad der bliver sagt om en given term.

- Specifik søgeterm
- Hvilken type af tweets, der ønskes vist (seneste¹³, mixet¹⁴ eller populære¹⁵)
- Antallet af tweets, der ønskes returneret

Via <http://beta.tagster.dk/latest/eurovision/mixed/15/>, søges der efter et *mix* af 15 tweets, hvor termet *eurovision* indgår. Resultatet af dette er illustreret i figur 9.5.

De returnerede tweets er præsenteret i samme stil, som det kendes fra Pinterest[57], hvor pladsen udnyttes bedst muligt. Dette layout går igen, når de 15 tweets skal præsenteres bedst muligt jf. pladsen på skærmen, så at der er størst fokus på indholdet. Hver tweet er visualiseret i hver sin boks. Disse er farvekodet, alt efter om der er tale om en positiv, negativ eller neutral stemt tweet. I headeren

¹³returnerer kun de seneste tweets.

¹⁴returnerer et mix af seneste og mest populære tweets.

¹⁵returnerer kun de mest populære tweets.

The screenshot shows a web application interface for Tagster.dk. At the top, there are navigation links for 'Dashboard' and 'Log out'. The main heading is 'Tweets eurovision'. Below this, there is a filter bar that says 'Now showing the mixed tweets for eurovision.' The content is organized into a grid of tweet cards. Each card contains the following information: a timestamp (e.g., 'Eurovision at 23:05:02 - 30-05-14'), the user's name and profile picture, the tweet text (often including links), and a score. Below the score, there is a small comment: 'The tweet is neutral, (0.0)' or 'Rank for user is awesome!, (98.32)'. The scores vary significantly, with some being very high (near 100) and others being very low (near 0 or negative).

Figur 9.5: Tagster.dk - Udforsk - Eurovision

i hvert boks står brugerens navn samt dato skrevet. For at fremhæve teksten i tweeten, er denne illustreret som et citat. Herunder er brugerens score listet samt en kommentar ud fra den givne score. Hvis scorereren er lav, så vil der blive præsenteret en mindre optimistisk tekst og vice versa hvis brugeren har en høj score.

9.4.2 Evaluering og validering af web-applikation

Gennem forløbet har der været en direkte kommunikation med personerne bag 140tegn.dk, som i afsnit 9.3.3.1 viste sig at være de første interesserede. Som et led i *build, measure, learn*-modellen i afsnit 4.1.1, fik de udleveret første version af web-applikationen således at der kunne indhentes feedback på denne. I Bilag J er en gennemgang af deres eksisterende løsning, samt en evaluering af web-applikationen (*dashboard, brugerprofil og udforsk*).

140tegn.dk er en social virksomhed, som tilbyder services til større virksomheder, og har bl.a. stået for håndteringen af sociale medier under eurovision samt

VM i fodbold for DR¹⁶.

Følgende er gennemgang af web-applikationen ud fra deres feedback i Bilag J.

9.4.2.1 Gennemgang af Dashboard

Det blev her fundet, at dashboardets øverste fire felter gav et godt overblik, men at en eventuel *"brand-score"* ikke kunne uddybes videre og den blev derfor fundet overflødig. *"Brand-score"* blev sammenlignet med Falcon Social's *"engagement level"*, hvilket blot gav endnu en score at holde styr på. Ønsket var hellere, at man så Twitters eksisterende *"retweets"*, *"favorites"* og *"mentions"* anvendt, hvilket ville give et reelt overblik over engagement fremfor en *opdigtet score*. I forbindelse med *"brand-mood"* blev den præsenterede score irrelevant, men at farven og en smiley gav mening og var tilstrækkelig.

For at skabe et bedre overblik over tid, var en funktion til at *"gå tilbage i tiden"* ønsket. Denne funktion tilbyder deres eksisterende løsning, og anvendes til at sammenligne nuværende data med tidligere indhentede. Farve indikationen blev dog fundet som en god ide, da den gav overblik og eventuelt kunne bruges til at fremhæve vigtige tweets.

De fandt både *"call-to-action"* og *"tagcloud"* yderst relevante, men i tagcloud'en var en kategorisering af enten ord eller hashtag savnet.

9.4.2.2 Gennemgang af Brugerprofil

Her blev et ønske om bedre grafisk forskel på *followers*, *following* etc. fundet, og at den opstillede tagcloud blev spredt ud på hhv. både hashtags og hyppigst anvendt ord. Det blev fundet, at den opstillede tagcloud var bedre end den opstillet på *"dashboard"*.

9.4.2.3 Gennemgang af Udforsk

Denne del blev fundet som den mindst nyttige og gav mindst overblik. Alt for meget tekst og for lidt fokus på indhold. At vise både klout-score samt tilhørende tekst blev fundet overflødig, og man savnede inddeling efter eksempelvis klout-score.

¹⁶Danmarks Radio

9.4.2.4 Gennemgang af Prisstruktur

Det blev yderligere kommenteret, at det prismæssigt var sat for højt, da de ikke selv ville være klar til at betale kr. 499 for produktet uden de efterspurgte funktioner. De fandt også, at prisstrukturen for de tre produkter ikke hang sammen, og opfordrer til et prismæssigt lavere produkt med færre funktioner, samt et højere spring til det næste på hhv. pris og funktioner. Til større virksomheder anbefales det, at man skræddersyr produktet efter kundens behov og ligeså med prisen.

9.5 Evaluering af 5. iteration

I dette kapitel er der efter lean-startup principperne blevet udarbejdet to MVP'ere. Den ene i form af en landing-page, der skal forsøge at promovere produktet, og indhente værdifuld information, om produktet faktisk er efterspurgt på markedet. Den anden i form af en web-applikation, der danner grundlaget for præsentationen af det indhentede relevante data.

Både Dropbox og Mailbox er eksempler på, hvordan den gode landing-page sættes op. Her anvendes både illustrationer samt video til at fremhæve deres produkter, og gennem det konvertere besøgende til kunder. Der vil i dette projekt blive anvendt dele af de samme nøgleelementer.

Projektets landing-page er inddelt i blok-sektioner, som hver især bidrager med, at fange de besøgenes interesse, og til sammen danne rammerne for produktet. En videopræsentation i samme stil, som set hos hhv. Dropbox og Mailbox, viser hvad produktet indeholder, er blevet udarbejdet. Denne er blevet placeret under headeren, således at en eventuel køber hurtigt bliver fanget og præsenteret for mulighederne ved produktet visuelt. Dette sker på baggrund af analysen, at en video kan holde de besøgende på siden i længere tid, men også fordi de besøgende alt andet lige, vil finde det nemmere at få fremvist produktet, fremfor at skulle læse om det.

For at kunne bedømme, hvorvidt produktet faktisk er efterspurgt på markedet, er der anvendt *call-to-actions* knapper på siden. Der er blevet anvendt hhv. Google Analytics samt bitly[5] til, at indhente antal besøgende og antal klik fra hver cta-knap.

En twitter-profil blev i februar oprettet og gennem denne, er billeder og features blevet delt flittigt. I april måned kom budskabet om produktet ud, da en

bruger på Twitter søgte efter et produkt, som kunne indhente kvalitativ data fra Twitter. Her blev *tagster.dk* nævnt som værende en løsning på netop denne problemstilling. Dette beviser, at sådanne produkter er efterspurgt på markedet. Twitter-profilen er også anvendt som kanal af trafik til den opsatte landing-page. Denne havde i perioden næsten 400 besøgende, de opsatte *call-to-actions* fik indhentet data omkring de besøgende. Her viste det sig, at de besøgende overvejende var interesseret i at lytte til hvad der bliver talt om på de sociale medier. Det blev også vist, at de besøgende var interesserede i produktet illustreret på siden, og gerne ville have deres eget dashboard. At videoen er blevet afspillet af næsten 1/4 af alle besøgende illustrerer interessen i et sådan produkt, og at de gerne vil vide mere om hvad produktet kan.

Det indhentede data viser også, at det er få af de besøgende, som enten skriver sig op til et nyhedsbrev eller bestiller et af de opsatte produkter. At udarbejde et nyt layout til en landing-page vil kunne afsøge, hvorvidt ændring af opstillingen vil have effekt på, hvor mange af de besøgende, som skriver sig op etc.

Den udviklede web-applikation indeholder de mest nødvendige funktioner, og kan derfor præsentere de relevante tweets samt det analyserede data, som er indhentet, filtreret og sorteret. Ved at anvende forskellige frameworks til opbygning af web-applikationen, blev det gjort muligt hurtigt og nemt at kunne få udarbejdet en MVP, som man kan gå til markedet med.

De tre mest nødvendige funktioner, *dashboard*, *brugerprofil* og *udforsk* bidrager med et overblik over forskellige typer af data indhentet. Dashboardet giver et overblik over al relevant information omkring et givent brand. Brugerprofilen giver indblik i en specifik bruger, hvor tweets, score og andre relevante oplysninger er gjort tilgængelige. Den sidste bidrager med et overblik over en specifik term, som brugere anvender. Netop denne funktion giver mulighed for at få indsigt i, hvad andre brugere taler om ud fra det søgte term.

Den første version af web-applikationen blev af virksomheden *140tegn.dk* testet og evalueret. I evalueringen blev der fundet gode og mindre gode emner. De fandt præsentationen samt det grafiske indhold godt, men savnede mere direkte indhold fra Twitter samt muligheden at få et overblik over tid. Generelt set var der her plads til forbedringer, og med deres feedback, er det nu muligt at udarbejde en ny version. Her kan der med fordel A/B-testes og se hvorvidt om de nye funktioner, vil forbedre web-applikationen.

Hypoteserne opstillet for produktet har vist sig at være korrekte og at man kan gå videre med produktet. Det viste sig, at produkter som kan indhente og vise analyseret data fra sociale medier er eftertragtet, hvilket også blev understøttet af antallet af besøgende på landing-pagen. I videopræsentationen af produktet, blev de besøgende vist hvilke funktioner produktet tilbyder fra start. Det kan

hverken be- eller afkræftes, hvorvidt de viste funktioner var tilstrækkelige, men en ny udgave af landing-pagen og en A/B kan eventuelt forsøge at bekræfte dette.

9.6 Validering af forretningsmodel

I denne iteration blev produktet evalueret, og baseret på feedback på første version er følgende segmenter i forretningsmodellen blevet revideret.

Det blev ud fra feedback fra `140tegn.dk` fundet, at problemstillinger tidligere opstillet ikke helt afspejlede kundernes egentlige problemstillinger. Her blev problemstillinger som *at kunne gå tilbage i tiden*, *kunne sammenligne hashtags* og *kunne se mønstre i realtidsdata* fundet. Topsy blev fundet, som en eksisterende løsning til disse problemstillinger. Problemstillingerne samt eksisterende alternativer er tilføjet til problem-segmentet i figur 9.6.

På baggrund af de reviderede problemstillinger, kunne der udarbejdes mulige løsninger til disse. Tre yderligere metoder er blevet tilføjet til den reviderede forretningsmodel, således at problemstillinger kan håndteres i en ny version af produktet. Disse er tilføjet til løsningssegmentet i figur 9.6.

Efter evalueringen blev det klart, at de første kunder også ønsker en sammenligning af et eller flere emner. Dette blev klart ud fra problemstillingen. Kunde-segmentet i figur 9.6 er blevet revideret med nye oplysninger.

Et emne som blev berørt i evalueringen af produktet var prisstrukturen. Her blev det fundet, at den billigste løsning ikke var prisbevidst overfor i særdeleshed mindre virksomheder. Et billigere basisprodukt var efterspurgt. På baggrund af hhv. evalueringen fra `140tegn.dk` og statistikken fra de opsatte produkter på landing-pagen, beskrevet i afsnit 9.3.3.1, er de nuværende produkter ændret til et standard "godt klædt på"produkt samt et "skræddersyet", der alt efter behov sammensættes.

Tagster - Forretningsmodel

<p>PROBLEM</p> <p>At indhente information omkring et specifikt emne</p> <p>At indhente kun relevant data og få frasorteret unødvendig data</p> <p>At indhente holdninger og meninger ud fra et brand</p> <p>At kunne gå tilbage i tiden</p> <p>At kunne sammenligne hashtags mod hinanden</p> <p>At se mønstre i realtidsdata</p> <p>EXISTING ALTERNATIVES</p> <p>BrandWatch</p> <p>Overskrift.dk</p> <p>Twitter</p> <p>HootSuite</p> <p>Falcon Social</p> <p>Topsy</p>	<p>SOLUTION</p> <p>Et emne-filter, som kan indhente data ud fra et emne</p> <p>Et spam-filter, som kan fjerne al ikke relevant data</p> <p>En metode, således meninger og holdninger kan udtrækkes</p> <p>En metode, der ud fra en specifik dato eller periode, kan indhente data</p> <p>En metode, således data ud fra bestemte hashtags kan sammenlignes</p> <p>En metode, således realtidsdata bliver overvåget og kan spotte eventuelle mønstre</p> <p>KEY METRICS</p> <p>Hovedområde: Indhente, analysere/filtere og præsentere social data</p> <p> Succeskriterie: Opbygge en virksomhed med 10 faste kunder efter første år</p>	<p>UNIQUE VALUE PROPOSITION</p> <p>Gør hvad I gør bedst, og lad os fortælle hvad folk mener om Jer</p> <p>Inet behov for søgeord eller svære forespørgsler, det hele sker automatisk</p> <p>Vi gør social data spiseligt</p> <p>HIGH-LEVEL CONCEPT</p> <p>Social data analyse uden al det unødvendige</p> <p>Social data analyse gjort simpelt</p>	<p>UNFAIR ADVANTAGE</p>	<p>CUSTOMER SEGMENTS</p> <p>#1 Mindre webshops</p> <p>#2 Marketingsbureauer</p> <p>#3 Nyhedsbureauer</p> <p>EARLY ADOPTERS</p> <p>Søger indblik i social medie omtale</p> <p>Ønsker en dybere forståelse af hvad folk mener om specifikke emner</p> <p>Ønsker sammenligningsgrundlag mellem to eller flere emner</p>
<p>COST STRUCTURE</p> <p>Hosting omk. På nuværende tidspunkt gratis lokal server.</p> <p>Udviklings timer er på nuværende tidspunkt gratis. Forventes at tilbyde partnerskab til eventuelle nye udviklere.</p> <p>AdWords kampagner. Afhænger af CTR. Budget sat lavt til 20kr/dag.</p> <p>Fremtidig hosting hos Google Cloud: Mindst kr. 212,39/md. Variabel omk. forbundet med mængde af data indhentet</p>		<p>REVENUE STREAMS</p> <p>#1 "Godt klædt på" med 14-dages prøveperiode uden binding, herefter 249kr/md. Begrenset på features</p> <p>#2 "Skreddersyet" uden prøveperiode. Pris afhænger af features og datamængde</p>		

Figur 9.6: Forretningsmodel for femte iteration

6. iteration

Sjette iteration vil gennemgå en videre analyse af, hvad der gennem social data kan beskrives. Her vil to cases blive anvendt, og gennem opstillede hypoteser vil data omkring *voteman* og *eurovision* blive analyseret, således at mønstre i det indhentede data vil kunne findes.

10.1 Social data analyse

Der er i dette projekt taget udgangspunkt i social data analyse ud fra to typer af cases. Følgende er derfor en gennemgang af, hvorfor de er valgt samt det ønskede resultat for hver af dem.

Voteman

Voteman er valgt på baggrund af dens kontroversielle indhold. Indholdet skabte forsider verden over, og på baggrund af klager til Folketinget blev den herefter trukket retur.

Det vil i forbindelse med projektet forsøges at gennemgå *Voteman* udelukkende på baggrund af social data. Det er her antaget, at det ud fra analyser er muligt at finde mønstre og eventuelt se, hvorvidt *Voteman* har haft en direkte indflydelse på deltagerprocenten ved Europaparlamentsvalget.

Eurovision

Eurovision er valgt på baggrund af dets store samtaleemne, anvendelsen af sociale medier under begivenheden og fordi brugerne har en indflydelse på udfaldet af vinderen.

I forbindelse med et tidligere kursus ved DTU¹, har man forsøgt at forudsige vinderen af Eurovision. I forbindelse med dette projekt[12] blev der indhentet historisk data fra 1998 til 2012. Det indhentede data blev analyseret og vha. statistiske modeller blev syv lande fra top-ti listen fundet. Da der udelukkende blev anvendt historisk data, var det ikke muligt at tage højde for påvirkningen fra sociale medier.

I dette projekt vil der forsøges at gennemgå Eurovision udelukkende på baggrund af social data under Eurovision. Det er antaget, at man ud fra analyser af social data, vil kunne finde mønstre og eventuelt finde vinderen af årets Eurovision, endda før finalen.

10.1.1 Voteman

I maj måned blev der afholdt Europaparlamentsvalg, og i den forbindelse blev der udarbejdet en valg-video, som skulle få de unge mennesker i stemmeboksene. I denne 90-sekunders video omhandlende *Voteman*, en muskuløs mand, der gennem vold skal få sofavælgerne ud af deres hjem og ned i stemmeboksene.

Videoen blev første gang lanceret på de sociale medier d. 12. maj, og her skabte den en mediestorm uden lige. Mange politikere ønskede videoen trukket retur pga. det kontroversielle indhold, og efter stærk kritik valgte Folketinget at trække videoen retur dagen efter, hvorefter den officielle video på YouTube blev gjort privat og dermed utilgængelig.

I det følgende afsnit vil der blive opstillede hypoteser for at afsøge mønstre, og hvorvidt det er muligt at beskrive forløbet vha. social data.

10.1.1.1 Hypoteser

Social data antages at kunne beskrive forskellige mønstre og gennem de følgende opstillede hypoteser, vil det på baggrund af indhentet data omkring *#voteman* blive undersøgt, om det faktisk er muligt at spotte disse mønstre.

- Det antages, at det ud fra indhentet data er muligt at spotte begivenheder i forbindelse med valg-videoen. Herunder hvornår den blev lanceret, hvornår den blev trukket retur og om det er muligt at spotte valgdagen.

¹Danmarks Tekniske Universitet

- Det antages, at det ud fra indhentet data er muligt at følge spredningen på tværs af lande globalt set.
- Det antages, at det ud fra indhentet data er muligt at bestemme hvorvidt videoen har haft en effekt på valgprocenten.

Følgende afsnit vil være en gennemgang af hvordan data er indhentet, samt resultater og evaluering af de opstillede hypoteser.

10.1.1.2 Indhentning af data

Et nødvendigt og komplet datasæt er nødvendigt, hvis mønstre ud fra opsatte hypoteser skulle findes. Idet ældre data også skulle undersøges var det nødvendigt at anvende opsatte søgninger vha. Twitters avanceret søgning. Her blev både *voteman* og *#voteman* anvendt som specifikke søgetermer og gennem følgende opstillede forespørgsler, blev det muligt at få et komplet datasæt² for *Voteman*.

- *#voteman OR voteman until:2014-05-13*³
- *#voteman OR voteman since:2014-05-13 until:2014-05-25*⁴
- *#voteman OR voteman since:2014-05-25*⁵

Disse tre forespørgsler gjorde det muligt først, at indhente samtlige tweets op til datoen⁶, hvor den omdiskuterede video gik viralt, herefter tweets i hele perioden op til valgdagen, samt til sidst efter valgdagen.

I tabel 2 i Bilag K er det illustreret, at der i perioden d. 12-05-2014 til d. 22-06-2014, er blevet indhentet næsten 6.000⁷ tweets, der er fordelt ud over de respektive datoer.

Det indhentede data er klar til videre analyse, og i følgende afsnit vil hypoteserne forsøges testet.

²Datasæt findes på <http://u.uhrebrink.dk/1qsL6Pi>

³<http://u.uhrebrink.dk/1p5KDFx>

⁴<http://u.uhrebrink.dk/UvnezH>

⁵<http://u.uhrebrink.dk/1111y3W>

⁶13-05-2014

⁷5.672

10.1.1.3 Hypotesetest

For at få et overblik over det indhentede data, er der blevet udarbejdet en wordcloud for de mest anvendte ord. I denne er de mest anvendte stopord fjernet, herunder de mest fremtrædende som *voteman* og *youtube*, således at kun relevant data fremvises. Denne wordcloud er illustreret i figur 10.1.



Figur 10.1: Mest anvendte termer for *#voteman*

De mest hyppige termer såsom *cartoon*, *video*, *denmark*, *folketinget*, *young* og *vote*, beskriver både videoen, dens afsender, samt hvad formålet er med den. Det ses desuden, at mange af termerne er på både spansk, italiensk og engelsk, hvilket understøtter hypotesen om en global omtale, der senere i dette afsnit vil blive undersøgt.

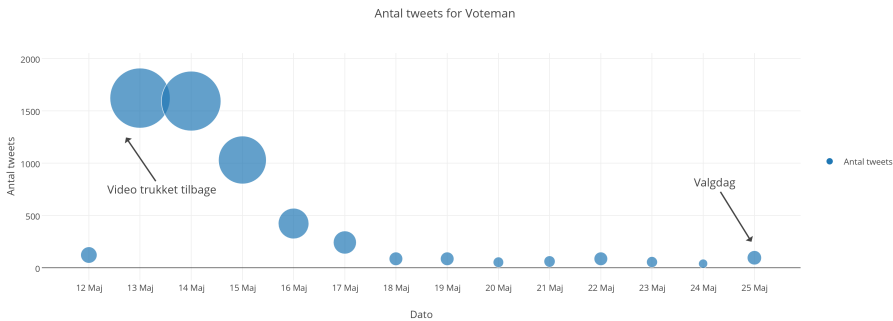
Mønstre i data

For at teste hypotesen, hvorvidt det er muligt at finde mønstre og begivenheder i det indhentede data, er data fra tabel 2 i Bilag K blevet analyseret, hvilket er illustreret i figur 10.2, samt yderligere i figur 34 i Bilag K.

Der kan ud fra figur 10.2 ses tydelige mønstre, og ud fra de eksisterende oplysninger omkring videoen og valgdagen, er der på grafen anvist hvornår videoen blev trukket retur, samt valgdagen for Europaparlamentsvalget. Det indhentede datasæt indeholder yderligere information om hvornår og hvem, der første gang nævnte *#voteman* med en reference til videoen⁸. Den første tweet er illustreret i figur 10.3.

De første par timer kommer videoen ud til et bredt publikum og især politikere^[60] har en stærk holdning til den, og videoen bliver et debateme i nyhederne samme aften d. 12. maj, hvor man ønsker videoen trukket retur. Den mængde omtale

⁸<http://u.uhrebrink.dk/1wiw1RZ>, som nu er trukket tilbage



Figur 10.2: Antal tweets for *#voteman* pr. dag. Før, under og efter valget



Dan Nielsen @DNNielsen · May 12
Folketingets valgfilm: **VOTEMAN** - bit.ly/1nFlvBg Meget meget meget mærkeligt... #dkpol

Figur 10.3: Første tweet om *#voteman*

videoen får, bliver på de sociale medier forstærket, da man d. 13. maj trækker videoen retur[59]. Her ses udslaget af mediestormen tydeligt på figur 10.2, hvor man går fra ca. 121 tweets om *#voteman* d. 12. maj til at have over 1600⁹ tweets den efterfølgende dag. Disse data er illustreret i tabel 2 i Bilag K.

Selvom Folketinget har trukket videoen retur, så kan man ikke *slukke internettet*, for når det først er gået viralt, så er det svært at stoppe internettets gang. I datasættet findes der 1.988¹⁰ links til videoer på YouTube, og på nuværende tidspunkt findes der over 5.000 udgaver af videoen om *voteman*, hvilket er illustreret i figur 36 i Bilag K.

Frem til valgdagen ses en tendens til færre personer, der på twitter omtaler *#voteman*, og på selve valgdagen er der igen, et forhøjet antal nye tweets, hvilket også er illustreret i figur 10.2. Efter valgdagen forsvinder *#voteman* næsten fra de sociale medier, idet den ikke længere er relevant, og derfor ses der kun få tweets, hvor *#voteman* er nævnt.

Spredning globalt set

For at teste hypotesen, hvorvidt det er muligt at følge den globale spredning for *#voteman*, er der ud fra hver tweet i det indhentede data¹¹, fundet det

⁹1622 tweets

¹⁰758 youtube.com samt 1.230youtu.be links

¹¹Datasæt findes på <http://u.uhrebrink.dk/1wn7WK0>

specifikke sprog. Twitter bestemmer allerede sproget for hver tweet, men ikke alle tweets indeholder lokationsoplysninger. Det har derfor ikke været muligt, at finde lokationen for hver tweet. Det er derfor antaget i dette projekt, at sproget også kan beskrive lokationen for den pågældende tweet. Ud fra denne antagelse er landet for hver tweet fundet, hvilket over tid vil kunne illustrere spredningen.

I figur 10.4 er spredningen vist i fire eksempler. Øverst til venstre, er landene for de første par timer d. 12. maj illustreret. Øverst til højre, er landene for d. 12. maj om aftenen illustreret. Nederst til venstre, er lande for den 13. maj illustreret, og til højre er det totale antal lande, hvor *#voteman* er blevet nævnt i perioden.



Figur 10.4: Global spredning af *#voteman*

Det blev i figur 10.1 illustreret de hyppigst anvendte termer ud fra det indhentede data. Her blev der fundet tweets med forskellige nationaliteter, hvilket nu over tid er blevet illustreret i figur 10.4. *#Voteman* var så kontroversiel, at den hurtigt ramte forsider rundt omkring i verden, hvilket også afspejler sig i det illustrerede data i figur 10.4. Det er hermed påvist, at man gennem social data kan følge udviklingen for et term eller et hashtag, som vist med *#voteman*.

Effekt på stemmeprocenten

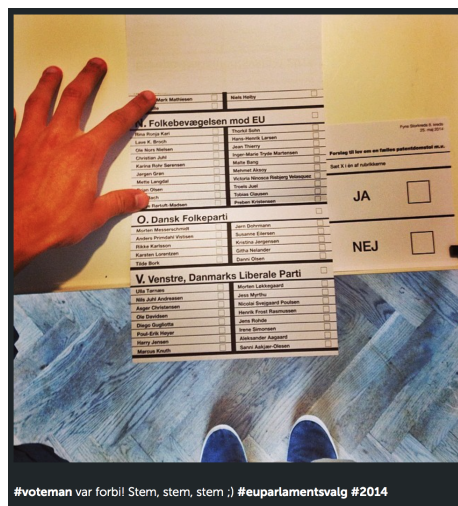
For at undersøge hvorvidt *#voteman* har haft en direkte indflydelse på stemmeprocenten ved årets Europaparlamentsvalg, er der ud fra det indhentede data undersøgt, hvorvidt det officielle twitter hashtag, illustreret i figur 10.5, optræ-

der sammen med enten *voteman* eller *#voteman*. Hvis disse termer optræder sammen, så må det antages, at *#voteman* har haft en indflydelse og dermed også en effekt på stemmeprocenten.



Figur 10.5: Det officielle *#jegstemmer* hashtag ved Europaparlamentsvalg

I det indhentede data, var der ikke belæg for påstanden om, at *#voteman* skulle have haft en direkte effekt på stemmeprocenten. Der blev i stedet søgt på det sociale medie Instagram, for her at undersøge billeder hvor *#voteman* og *#jegstemmer* optræder sammen.



Figur 10.6: *#Valgfie* fra Instagram, hvor *#voteman* er anvendt

Det var ikke muligt, at finde billeder hvor begge hashtags var anvendt. Dog viste det sig, at *#voteman* ofte blev anvendt på såkaldte *#valgfies*, hvor man har taget et billede af sig selv i stemmeboksen, hvilket er illustreret i figur 10.6.

Hvorvidt *#voteman* har haft en direkte indflydelse på stemmeprocenten kan være svært at måle, men at mange personer anvender hashtagget i forbindelse med billeder taget fra stemmeboksene, der er illustreret i Bilag L, kan vise tegn

på en sammenhæng. Yderligere blev personer på hhv. Instagram og Twitter udspurgt om, hvorvidt #voteman fik dem til at stemme, og hvad deres holdning til videoen var. Kommentarerne er illustreret i figur 39 og figur 40 i Bilag M. Videoen har skabt omtale på hhv. Instagram og Twitter, men om den har haft en egentlig effekt på valgprocenten, kan dog hverken be- eller afkræftes.

10.1.2 Eurovision

I starten af maj måned blev der i traditionen tro, afholdt det årlige Melodi Grand Prix. Da Danmark vandt sidste år, fik Danmark æren af at være vært for showet i år. Eurovision er blevet målt som en af verdens største årlige tv begivenheder i verden[20], og kan derfor være svært at komme helt udenom.

Siden 2008[85] har showet foregået over tre dage. To semifinaler og en finale. I år fandt semifinalerne sted *tirsdag d. 6. maj* og *torsdag d. 8. maj*, hvor finalen fandt sted *lørdag d. 10.* Der er derfor tid inden finalen, hvor publikum kan nå at høre de forskellige sange og danne sig en mening. De sociale medier er, alt andet lige, steder hvor folk kan dele deres meninger om landenes sange.

10.1.2.1 Hypoteser

Da der bliver skrevet meget på de sociale medier inden, under og efter de tre eurovision live udsendelser, vil det forsøges ved hjælp af opstillet hypoteser, at spotte mønstre og trends.

- Det antages, at det er muligt at spotte forskellige begivenheder i forbindelse med Eurovision. Herunder hvornår live showet starter, hvornår særlige vigtige kunstnere optræder og hvornår de går af scenen.
- Det antages, at det er muligt at forudsige vinderen inden finalen

I følgende afsnit vil der blive gennemgået hvordan data er indhentet, hvad det viser samt evaluering af de opstillede hypoteser.

10.1.2.2 Indhentning af data

På samme måde som i afsnit 10.1.1, blev der brugt Twitters avanceret søgning fra ugen op til, samt under forløbet. Der blev søgt på specifikke termer som var

eurovision, *#eurovision*, *joinus* og *#joinus* og ved hjælp af følgende forespørgsel, blev det muligt at få et komplet datasæt for ugen hvor Eurovision showet fandt sted.

- *eurovision OR #eurovision OR joinus OR #joinus since 2014-05-04 until: 2014-05-09*¹²
- *eurovision OR #eurovision OR joinus OR #joinus since 2014-05-11*¹³

De to forespørgsler har gjort det muligt, at skabe et overblik over ugen hvorpå Eurovision fandt sted, samt efterfølgende at se hvordan diverse emner bliver diskuteret. Dette har givet et datasæt på 2913 unikke tweets i løbet af ugen, hvor der ikke er tale om retweets, hvid støj eller andre irrelevante tweets.

Der er derudover blevet hentet danske tweets på aftenen af den første semifinale, med et omfattende datasæt¹⁴ på 2825 danske tweets. Dette vil kunne bruges til at følge udviklingen for den første aften, samt vurdere stemningen hos de forskellige lande.

I tabel 10.1 er det vist hvor mange danske tweets, der blev hentet under den første semifinale. De fleste tweets er centreret mellem klokken 21 og 23, hvilket stemmer med, hvornår showet fandt sted.

Timer	Antal
20:00 - 21:00	76
21:00 - 22:00	1476
22:00 - 23:00	1204
23:00 - 24:00	67

Tabel 10.1: Antal tweets i forskellige tidsrum

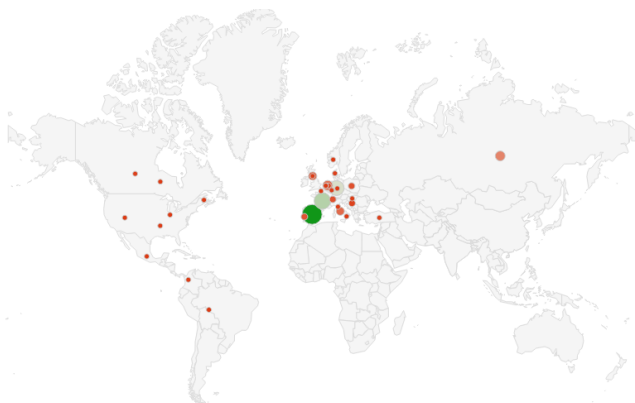
10.1.2.3 Hypotese test

For at få et overblik over hvilke lande der taler om Eurovision, er mængden af tweets pr. land blevet gemt. Resultatet af dette ses i figur 10.7, som viser det meste af verdenen lytter med. Den største del af tweets er dog centreret omkring Europa, da det er disse lande som deltager.

¹²<http://bit.ly/T2E87v>

¹³<http://bit.ly/1iyEdf3>

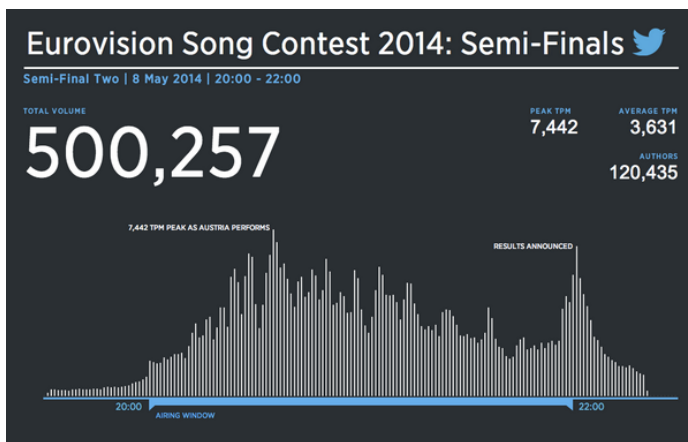
¹⁴Datasæt findes på <http://u.uhrebrink.dk/UAqj0T>



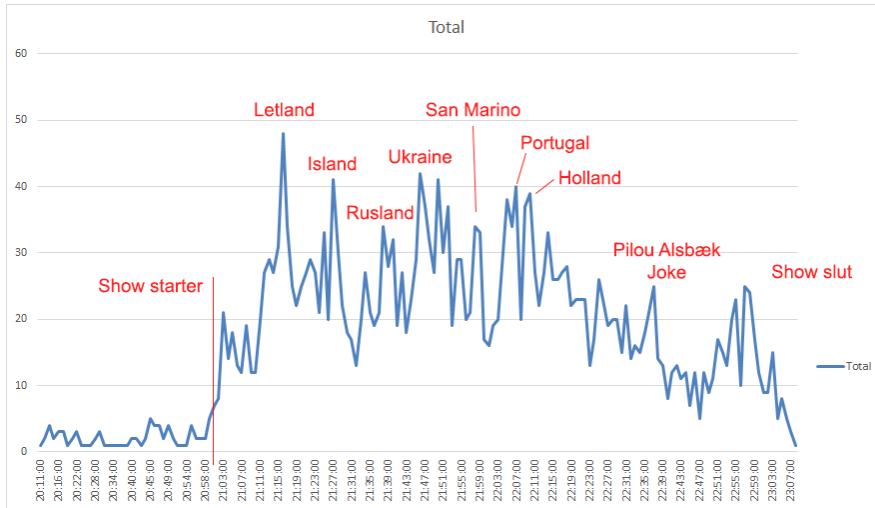
Figur 10.7: Spredning af lande der taler om Eurovision

For at teste om hypoteserne holder, bliver data analyseret yderligere.

Mønstre i data Den første semifinale fandt sted tirsdag den 6. maj mellem kl. 21 til lidt efter 23 hvilket også kan ses udfra mængden af danske tweets i tidsrummet i tabel 10.1. I løbet af de to timer blev der sendt utrolig mange tweets vedrørende Eurovision. På Eurovisions egen Twitter side[25] vises en statistik, over hvor mange der var tale om i alt, samt forløbet under showet som også er illustreret i figur 10.8. Som det ses, er der tale om en halv million tweets i løbet af lidt over to timer. Dette viser sociale medier er glimrende til at dele lignende begivenheder.



Figur 10.8: Spredning af danske tweet i løbet af den første semifinale



Figur 10.9: Spredning af danske tweet i løbet af den første semifinal

Ud fra det danske datasæt som blev gemt undervejs, ses nogenlunde samme trends. Figur 10.9 viser danske tweets og forløbet undervejs. Det ses, at showet starter kl. 21 med antallet af tweets begynder at stige, samt hvilke sange der vækker særlig opmærksomhed. I toppen ligger Letlands "Cake to bake" som gav en masse omtale fra de kom på kl. 21:15 og aftenen ud. De fleste tweets vedrørende Letland mindede alle om dem vist i tabel 10.2, og var ikke decideret positive om Letlands chancer for at vinde.

Tid

Tweets

21:15:04 *Letlands sang er lige i Brødrene Price og Mette Blomsterbergs smag*

21:17:39 *Letland: En hash-kage - det skal jeg i hvert fald have, hvis den går videre*

21:18:01 *Letland - Seriøst!! Hvor dårlige har de andre sange så været!*

Tabel 10.2: 3 tweets der handler om Letlands "Cake to bake"

Selvom showet peakede med det højeste antal danske tweets kl. 21:15, hvilket er illustreret i figur 10.9, med flest omtaler vedrørende Letlands sang fortsatte antallet af tweets stadig støt undervejs. Ved at se på hvor mange gange de forskellige lande bliver nævnt undervejs, var det muligt at se forskellige trends i løbet af showet. Tabel 10.3 viser placeringen af de syv lande, som modtog flest point samt antallet af danske tweets under showet.

Som det ses viser antallet af tweets ikke en klar vinder af den første semifinal.

Placering	Antal tweets	Lande
1	83	Netherlands
2	63	Sweden
3	37	Hungary
4	26	Armenia
5	100	Ukraine
6	94	Russia
7	33	Montenegro

Tabel 10.3: Placering til den første semifinale kontra antal tweets

Det ses dog, at både Ukraine og Rusland er blevet omtalt en del. Dette skyldes primært konflikten mellem disse to lande, hvor tweets ikke ligefrem er positive. Eksempler på sådan to tweets er illustreret i tabel 10.4 som viser, at ikke al omtale er godt.

Tid	Tweets
21:40:16	<i>Skal man stemme på Rusland for ikke at blive invaderet af Putin?</i>
21:24:10	<i>Som noget nyt i år vil halvdelen af Ukraines stemmer automatisk blive overtaget af Rusland.</i>

Tabel 10.4: Tweets som nævner Rusland

Hvis der ses bort fra Ukraine-Rusland konflikten, ses det i tabel 10.3, at der er et sammenhæng mellem omtale og hvilke lande der fik flest point.

Det er derudover også muligt at se hvornår hvilke lande synger. Et eksempel på dette er San Marino. Dette er for mange et ukendt land, som kun bliver nævnt i forbindelse med Eurovision. Nok også derfor skaber dette også en del omtale omkring deres optræden hvor alle 48 tweets som nævner San Marino finder sted mellem 21:57 og 22:01, som passer meget godt med hvornår de optrådte. Eksempler på tweets om San Marino ses i figur 10.5 som også tydeligt viser interesse, men ikke troen på de at kan vinde¹⁵.

Tid	Tweets
21:57:49	<i>Plejer San Marino at være med?</i>
19:58:57	<i>Jeg elsker at San Marino altid er med; landet som ingen kender - og alle glemmer efter</i>

Tabel 10.5: Tweets som nævner San Marino's håb om at vinde Eurovision

¹⁵San Marino kom videre til finalen, men endte på en næstsidste plads[85]

Kort efter optræder Holland med den sang som endte med at ligge i toppen af denne semifinale. Med at blive nævnt 83 gange og primært omkring tidsrummet 22:08 til 20:10, stemmer det med at de optrådte næsten lige efter San Marino. I tabel 10.6 er illustreret eksempler på positive tweets samt troen på et hit¹⁶.

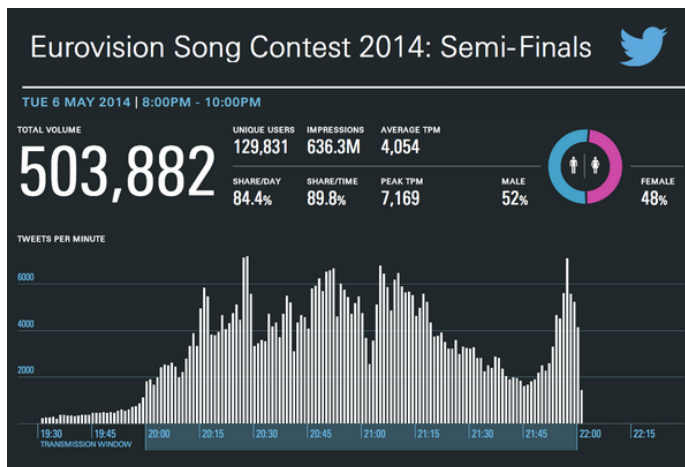
Tid	Tweets
22:09:24	<i>Hollands er ikke helt dum!! Ala stille raveonettes</i>
22:09:37	<i>Holland vinder, over and out!</i>

Tabel 10.6: Tweets der roser Holland

Det ses herved at det er muligt at følge trends og begivenheder i løbet af et show som Eurovision. Grundet antal tweets samt hvad der bliver nævnt i disse, er det muligt at se hvornår de forskellige lande optræder, samt hvad der tales om i tweets. Eksemplet med Ukraine og Rusland indikerer samtidig at antallet af tweets ikke er alt sigende, og det også er vigtigt at se, hvad der faktisk står i de tweets.

Hvem vinder

Under anden semifinale optrådte Conchita Wurst på vegne af Østrig. Hvorvidt man var tilhænger eller ej, var det svært at komme udenom at "*The bearded lady*" tog anden semifinale med storm[33]. Figur 10.10 er taget fra Eurovisions Twitter side[25] og viser hvordan antallet af tweets nåede højdepunktet mens Østrig sang.



Figur 10.10: Twitter peaker under Østrigs optræden

¹⁶Holland endte med at blive nummer to til finalen[85]

Ved at se på dagene op til finalen om lørdagen er det muligt et se trends begynde at starte samt stemningen vedrørende betydningsfulde emner. Ud fra det datasæt blev det klart, hvilke lande der var omtalt mest. Ved at finde de mest nævnte blev det muligt at lave følgende wordcloud som er illustreret i figur 10.11.



Figur 10.11: Top brugte ord under Eurovision ugen udover *Eurovision og Jo-inus*

Den illustrerer hvad der er blevet nævnt i løbet af ugen, hvor tendenser kan ses. *Rusland, Holland, Østrig* og især *Conchita Wurst* er klar tydelige. Ved at kigge nærmere på data blev det tydeligt at to lande som blev nævnt mest var Rusland og Østrig. Som tabel 10.7 viser er Rusland blevet nævnt flest gange, også mere end Østrig. Det er dog også tydeligt ud fra deres stemningsanalyse, som er beregnet på samme måde som i afsnit 8.3, at omtalen om Rusland er meget skiftende.

Land	Antal	Sentiment
Rusland	195	0.65
Østrig	119	3.04

Tabel 10.7: Rusland slår med antal af tweets, selvom stemningen er mere positiv for Østrig

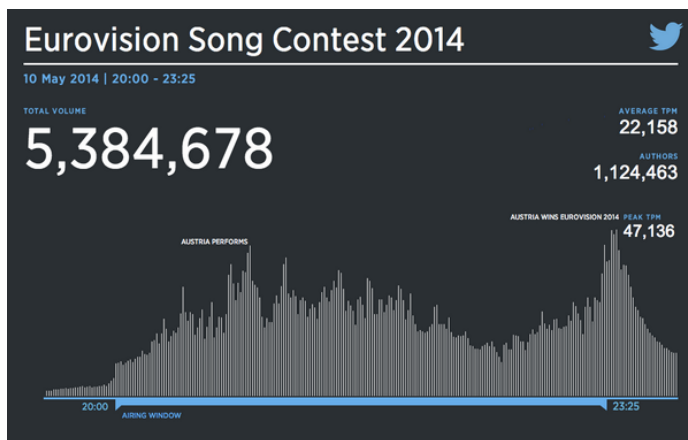
Da der er tale om næsten 200 tweets, som nævner Rusland specifikt og scoren er tæt på neutral, er der tegn på blandet omtale. I hvert fald ikke overvejende positiv. Ved at sammenligne med Østrig, er omtalen allerede meget mere positiv selvom antallet af tweets ikke er så højt i forhold til Rusland. Eksempler på hvordan Østrigs tweets er mere positive ses i tabel 10.8

Ligeledes kan det ses under selve finalen, at antallet af tweets når Østrig optræder er klart højest, som vist i figur 10.12. Ud fra stemningsanalysen og antallet af

Tid	Tweets
2014-05-08 17:21:37	<i>If @ConchitaWurst does not win Eurovision this year something is wrong. By far the stand out performance of semis</i>
2014-05-08 13:26:50	<i>Austria's Conchita Wurst fends off transphobia before the Eurovision Song Contest</i>
2014-05-08 14:35:38	<i>Just hear those cheers for Conchita!</i>
2014-05-08 14:35:24	<i>So powerful! The best of Conchita's performances. Love it!</i>

Tabel 10.8: Tweets som viser støtte til Østrigs sang

tweets omkring landene, er Østrig vurderet som værende vinderen af Eurovision, hvilket herved bekræfter hypotesen.



Figur 10.12: Conchita løber af sted med sejren og antallet af tweets

10.2 Evaluering af 6. iteration

I dette kapitel er der ud fra to cases, *voteman* og *eurovision*, blevet opstillede hypoteser for undersøgelse af mønstre i social data. *Voteman* blev valgt på baggrund af dens kontroversielle indhold og sociale omtale. *Eurovision* blev valgt ud fra dens samtale emne, anvendelsen af sociale medier og at brugerne havde en indflydelse på udfaldet.

Omkring *Voteman* blev der opstillet tre hypoteser for at kunne spotte begivenheder i forbindelse med videoen, følge spredningen globalt set og hvorvidt videoen har haft en effekt på valgprocenten.

Ved at gøre brug af Twitter avanceret søgning blev der vha. specifikke søgetermer indhentet et komplet datasæt. Ud fra dette datasæt blev videre analyse foretaget.

Antallet af tweets pr. dag i perioden blev visualiseret og gennem dette kunne begivenheder som hvornår videoen blev lanceret, hvornår den blev trukket tilbage og tilmed valgdagen blive illustreret.

For at følge spredningen globalt set, blev det antaget at sproget i en given tweet kunne beskrive dets oprindelse. Ud fra denne antagelse kunne spredningen for voteman over tid blive fastlagt.

Det blev antaget, at hvis *#voteman* og det officielle hashtag omkring Europaparlamentsvalget optrådte sammen, ville voteman have haft en effekt på stemmeprocenten. Dette kunne ikke understøttes og der blev i stedet anvendt Instagram til sammenligning. Her kunne man ud fra såkaldte *valgfies* se en sammenhæng, men en direkte effekt kunne hverken be- eller afkræftes herom.

Omkring Eurovision blev der ligeledes opstillet hypoteser for at se, hvornår de forskellige lande går på og andre begivenheder undervejs, samt hvorvidt det er muligt ud fra social data at bestemme vinderen af årets Eurovision, inden de går på scenen.

Der blev også anvendt Twitter avanceret søgning, således at et komplet datasæt før, under og efter eurovision kunne indhentes. Yderligere blev der under første semifinale indhentet danske tweets.

Der blev ud fra det danske datasæt muligt at se hvornår showet startede og sluttede, og yderligere se hvornår de enkelte lande optrådte. Undervejs blev andre begivenheder som jokes og debatten mellem Ukraine og Rusland også fundet.

Ud fra det indhentede datasæt var det muligt, at få et overblik over de mest omtalte lande på de sociale medier. Her blev Rusland og Østrig fundet som de to lande med størst omtale. En stemningsanalyse kunne herefter bekræfte, at Østrig havde den mest positive omtale, og blev vurderet som vinderen af eurovision allerede inden finalen.

10.3 Validering af forretningsmodel

Som beskrevet i forretningsmodellen i afsnit 4.2.1.9 forblev segmentet for urimelige fordele tomt, for hele tiden at kunne tilstræbe funktioner, der vil kunne give urimelige fordele på markedet.

Det blev i denne iteration forsøgt testet at kunne spotte mønstre i social data og ud fra disse eventuelt kunne forudsige hændelser. Gennem to cases blev dette påvist at være muligt. Disse kan på nuværende tidspunkt anses som at være urimelige fordele, da ingen eksisterende løsninger fundet har samme funktionalitet. Den reviderede forretningsmodel er illustreret i figur 10.13.

Tagster - Forretningsmodel - første iteration				
PROBLEM At indhente information omkring et specifikt emne At indhente kun relevant data og få frasorteret unødvendig data At indhente holdninger og meninger ud fra et brand At kunne gå tilbage i tiden At kunne sammenligne hashags mod hinanden At se mønstre i realtidsdata EXISTING ALTERNATIVES BrandWatch Overskrift.dk Twitter HootSuite Falcon Social Topsy	SOLUTION Et emne-filter, som kan indhente data ud fra et emne Et spam-filter, som kan fjerne al ikke relevant data En metode, således meninger og holdninger kan udtrækkes En metode, der ud fra en specifik dato eller periode, kan indhente data En metode, således data ud fra bestemt hashags kan sammenlignes En metode, således realtidsdata bliver overvåget og kan spotte eventuelle mønstre KEY METRICS Hovedområde: Indhente, analysere/filtere og præsentere social data Succeskriterie: Opbygge en virksomhed med 10 faste kunder efter første år	UNIQUE VALUE PROPOSITION Gør hvad I gør bedst, og lad os fortælle hvad folk mener om Jer Intet behov for søgeord eller svære forespørgsler, det hele sker automatisk Vi gør social data spiseligt HIGH-LEVEL CONCEPT Social data analyse uden al det unødvendige Social data analyse gjort simpelt	UNFAIR ADVANTAGE Se mønstre i realtid At kunne forudsige hændelser ud fra mønstre i data CHANNELS Venner og bekendte Mund-til-mund metoden Sociale medier (viral markedsføring) herunder Facebook og Twitter Landing-page AdWords kampagner	CUSTOMER SEGMENTS #1 Mindre webshops #2 Marketingsbureauer #3 Nyhedsbureauer EARLY ADOPTERS Søger indblik i social medie omtale Ønsker en dybere forståelse af hvad folk mener om specifikke emner Ønsker sammenligningsgrundlag mellem to eller flere emner
COST STRUCTURE Hosting omk. På nuværende tidspunkt gratis lokal server. Udviklings timer er på nuværende tidspunkt gratis. Forventes at tilbyde partnerskab til eventuelle nye udviklere. AdWords kampagner. Afhænger af CTR. Budget sat lavt til 20kr/dag. Fremtidig hosting hos Google Cloud: Mindst kr. 212,39/md. Variabel omk. forbundet med mængde af data indhentet		REVENUE STREAMS #1 "Godt klædt på" med 14-dages prøveperiode uden binding, herefter 249kr/md. Begrænset på features #2 "Skreddersyet" uden prøveperiode. Pris afhænger af features og datamængde		

Figur 10.13: Forretningsmodel for seks iteration

Konklusion

I forbindelse med dette projekt er der udarbejdet 1. version af en web-applikation, hvis formål er at præsentere analyseret social data, som er relevant for virksomhederne. Ved at omdanne kvantitativ data til kvalitativ data, kan der skabes overblik på et ellers uoverskueligt marked.

Inden udarbejdelsen af produktet er der foretaget en undersøgelse af de største sociale netværk, således at et datagrundlag for produktet kunne bestemmes. Analysen viser, at den nødvendige mængde data er til rådighed, og at brugere på daglig basis aktivt bidrager med mere indhold. Der blev gennem en analyse af markedet fundet eksisterende alternativer, der kan give virksomheder overblik på sociale medier. Med udgangspunkt i markedsanalysens resultater er web-applikationen blevet udarbejdet, således at prisstruktur matcher markedet og gennem fortolkning af data kan differentiere sig fra de eksisterende alternativer.

På baggrund af et hurtigt bevægende marked, er projektet kørt som et lean-startup. En forretningsmodel er udarbejdet og gennem iterationer er den på baggrund af ny viden blevet evalueret og revideret, således at en mere præcis forretningsmodel kan skabes.

Ud fra løsninger til problemstillinger fundet i forretningsmodellen er hypoteser opstillet, der danner grundlag for produktet. Gennem en iterativ proces er data fra Twitter testet mod de opstillede hypoteser, og er fundet struktureret nok til videre analyse. Forskellige metoder for indhentning af data er testet, og en kombination af *sample* og *search* er fundet tilstrækkelige til projektet. Twitter indeholder store mængder data, og gennem brand-specifikke søgetermer er der indhentet målrettet data. En definition af *hvid støj* er udarbejdet, og ud fra denne er ikke-relevant data frasorteret, således at kun relevant data står tilbage. Ved anvendelsen af frekvensfordeling og forskellige metoder for stemningsanalyse er det muligt at finde beskrivende og relevante ord omkring et brand, samt at udtrække meninger og holdninger omkring et specifikt emne. For at rangere og

sortere på relevant data, er der udarbejdet en definition af både en relevant bruger og af en vigtig tweet. Dette gør det muligt at præsentere data efter relevans og spotte eventuelle *call-to-actions*, som man bør reagere på.

Der er i projektet blevet udarbejdet to *Minimum Viable Products*, hvor det ene danner grundlag for det analyserede og relevante data, og den anden forsøger at promovere produktet og undersøge, om der er efterspørgsel på dette. Den første, i form af en web-applikation, der er testet af et socialt mediebyrå i kundesegmentet. På baggrund af feedback og erfaringer om kunders behov, er forretningsmodellen revideret. Den anden, i form af en landing-page, hvor statistik fra besøgende viser tegn på efterspørgsel af produktet.

En social data analyse er udarbejdet, således at der gennem to cases kan findes mønstre i data, og dermed kunne differentiere sig yderligere fra eksisterende alternativer. Ud fra analysering af social data, er det muligt at spotte begivenheder i realtid. Begivenheder som da #Voteman videoen blev trukket retur, hvordan #Voteman over tid spredte sig globalt set, hvornår de enkelte lande gik på scenen under Eurovision, samt at kunne forudsige udfaldet af Eurovision inden finalen. På baggrund af denne ny viden omkring mønstre i data, bidrager et nyt produkt med urimelige fordele i forhold til de eksisterende alternativer, og vil dermed danne grundlag for videreudvikling af produktet.

Litteratur

- [1] T. Ahlqvist et al. Social media roadmaps. <http://www.vtt.fi/inf/pdf/tiedotteet/2008/T2454.pdf>, 2014. [Online; sidst set d. 18-05-2014].
- [2] P. P. Alexander Pak. Twitter as a corpus for sentiment analysis and opinion mining. [http://incc-tps.googlecode.com/svn/trunk/TPFinal/bibliografia/Pak%20and%20Paroubek%20\(2010\)%20Twitter%20as%20a%20Corpus%20for%20Sentiment%20Analysis%20and%20Opinion%20Mining.pdf](http://incc-tps.googlecode.com/svn/trunk/TPFinal/bibliografia/Pak%20and%20Paroubek%20(2010)%20Twitter%20as%20a%20Corpus%20for%20Sentiment%20Analysis%20and%20Opinion%20Mining.pdf), 2014. [Online; sidst set d. 17-06-2014].
- [3] Atcore. Infographic: Danske brugere på twitter. <http://atcore.dk/blog/infographic-danske-brugere-pa-twitter/>, 2014. [Online; sidst set d. 04-06-2014].
- [4] G. Berkeley. Github. <http://u.uhrebrink.dk/1noMnUP>, 1710. [Online; sidst set d. 20-06-2014].
- [5] bitly. Take control of your links. <http://bitly.com/>, 2014. [Online; sidst set d. 01-06-2014].
- [6] Bootstrap. The most popular front-end framework for developing responsive, mobile first projects on the web. <http://getbootstrap.com>, 2014. [Online; sidst set d. 18-05-2014].
- [7] Brandwatch. Konkurrent forside. <http://www.brandwatch.com/>, 2014. [Online; sidst set d. 06-06-2014].
- [8] S. M. Business. The ultimate guide to minimum viable products. <http://scalemybusiness.com/the-ultimate-guide-to-minimum-viable-products/>, 2014. [Online; sidst set d. 30-05-2014].
- [9] S. Bøgh-Andersen. Danske twitter-brugere fulgt på overskrift.dk. <https://twitter.com/nitoen/status/471336396385689600>, 2014. [Online; sidst set d. 07-06-2014].
- [10] D. Counsell. Paid, paymium or freemium. <http://dancounsell.com/articles/paid-paymium-or-freemium>, 2014. [Online; sidst set d. 06-06-2014].

- [11] Datasift. Datasift. <http://datasift.com/>, 2014. [Online; sidst set d. 04-06-2014].
- [12] B. D. H. o. K. M. C. David Kofoed Wind, Helge Munk Jacobsen. Analysing the eurovision song contest (2013). <http://bit.ly/1lmayFJ>, 2013. [Online; sidst set d. 23-06-2014].
- [13] T. Developer. Entities in twitter objects. <https://dev.twitter.com/docs/entities1>, 2014. [Online; sidst set d. 04-06-2014].
- [14] T. Developer. The streaming apis. <https://dev.twitter.com/docs/streaming-apis>, 2014. [Online; sidst set d. 04-06-2014].
- [15] T. developer site. Rest api v1.1 limits per window by resource. <https://dev.twitter.com/docs/rate-limiting/1.1/limits>, 2014. [Online; sidst set d. 23-06-2014].
- [16] T. Developers. Beskrivelse af restrictioner på twitters rest search api. <https://dev.twitter.com/docs/rate-limiting/1.1>, 2014. [Online; sidst set d. 19-05-2014].
- [17] U. Dictionary. Hashjacking. <http://www.urbandictionary.com/define.php?term=hashjacking>, 2014. [Online; sidst set d. 10-06-2014].
- [18] E. DIgital. Making video accountable. <http://www.eyeviewdigital.com/documents/EyeView-White-Paper-Making-Video-Accountable.pdf>, 2014. [Online; sidst set d. 30-05-2014].
- [19] N. . Documentation. nltk.tokenize package. <http://www.nltk.org/api/nltk.tokenize.html>, 2014. [Online; sidst set d. 16-06-2014].
- [20] DR. Hvilket tv-show er størst? <http://www.dr.dk/Nyheder/Kultur/2014/03/05/161221.htm>, 2014. [Online; sidst set d. 23-06-2014].
- [21] Dropbox. Backup your life. <https://dropbox.com>, 2014. [Online; sidst set d. 30-05-2014].
- [22] e guillotine. How social media changed the way we interact with the world. <http://www.e-guillotine.com/how-social-media-changed-the-way-we-interact-with-the-world.php>, 2014. [Online; sidst set d. 27-06-2014].
- [23] C. egg. Don't like klout? 12 other ways to track social media influence and engagement. <http://blog.crazyegg.com/2013/06/04/dont-like-klout/>, 2014. [Online; sidst set d. 18-06-2014].
- [24] B. R. Elberth. Twitter ændrer nyhedens anatomi. <http://www.elberth.dk/twitter-aendrer-nyhedens-anatomi.html>, 2013. [Online; sidst set d. 24-05-2014].

- [25] T. Eurovision. Eurovision billeder. <https://twitter.com/search?q=eurovision&src=typd&mode=photos>, 2014. [Online; sidst set d. 23-06-2014].
- [26] Facebook. Connect with friends and the world. <https://facebook.com>, 2014. [Online; sidst set d. 22-05-2014].
- [27] Falcondai.com. 666 and how twitter samples tweets in streaming api. <http://blog.falcondai.com/2013/06/666-and-how-twitter-samples-tweets-in.html>, 2014. [Online; sidst set d. 04-06-2014].
- [28] Falconsocial. Konkurrent forside. <http://www.falconsocial.com/>, 2014. [Online; sidst set d. 06-06-2014].
- [29] FirstMonday.org. Detecting spam in a twitter network. <http://firstmonday.org/article/view/2793/2431>, 2014. [Online; sidst set d. 12-06-2014].
- [30] D. W. Framework. The web framework for perfectionists (with deadlines). django makes it easier to build better web apps more quickly and with less code. <https://www.djangoproject.com/>, 2014. [Online; sidst set d. 01-06-2014].
- [31] Geckoboard. Konkurrent forside. <http://www.geckoboard.com/>, 2014. [Online; sidst set d. 06-06-2014].
- [32] Gnip. Gnip. <http://gnip.com/>, 2014. [Online; sidst set d. 04-06-2014].
- [33] good-evening europe.dk. Conchita wurst tog europa med storm. <http://good-evening-europe.dk/2014/05/11/conchita-wurst-tog-europa-med-storm/>, 2014. [Online; sidst set d. 23-06-2014].
- [34] Google. Google web analytics. <http://www.google.com/analytics/>, 2014. [Online; sidst set d. 12-06-2014].
- [35] Google+. Google's social network. <https://google.com/plus>, 2014. [Online; sidst set d. 22-05-2014].
- [36] Hootsuite. Konkurrent forside. <https://hootsuite.com/>, 2014. [Online; sidst set d. 06-06-2014].
- [37] B. Hughes. Total antal ja og nej mentions ved patentdomstol på twitter. <http://patentdomstol.herokuapp.com>, 2014. [Online; sidst set d. 22-05-2014].
- [38] B. Hughes. Twitter status. <https://twitter.com/DalsHughes/status/466875499642646528>, 2014. [Online; sidst set d. 22-05-2014].

- [39] IAG.Me. Do you know your twitter limit? <http://iag.me/socialmedia/guides/do-you-know-the-twitter-limits/>, 2014. [Online; sidst set d. 22-05-2014].
- [40] Instagram. Capture and share the world's moments. <https://instagram.com>, 2014. [Online; sidst set d. 22-05-2014].
- [41] Instagram. Instagram press. <http://www.nytimes.com/2014/02/15/technology/the-plus-in-google-plus-its-mostly-for-google.html>, 2014. [Online; sidst set d. 07-06-2014].
- [42] R. John. How to validate your idea with a wizard of oz mvp. <http://ramlijohn.com/westjet-christmas-how-to-validate-your-idea-with-a-wizard-of-oz-mvp/>, 2014. [Online; sidst set d. 30-05-2014].
- [43] T. W. S. Journal. When freemium fails. <http://online.wsj.com/news/articles/SB10000872396390443713704577603782317318996?mg=reno64-wsj>, 2014. [Online; sidst set d. 06-06-2014].
- [44] Klout. klout score. <http://klout.com/corp/score>, 2014. [Online; sidst set d. 27-06-2014].
- [45] LinkedIn. The next three billion. <http://blog.linkedin.com/2014/04/18/the-next-three-billion/>, 2014. [Online; sidst set d. 07-06-2014].
- [46] LinkedIn. Worlds largest professional network. <https://linkedin.com>, 2014. [Online; sidst set d. 22-05-2014].
- [47] Mailbox. Mailboxapp. <https://mailboxapp.com>, 2014. [Online; sidst set d. 30-05-2014].
- [48] Mashable. Twitter opkøber gnip for at kunne yde bedre service til udviklere. <http://mashable.com/2014/04/15/twitter-buys-gnip/>, 2014. [Online; sidst set d. 20-05-2014].
- [49] A. Maurya. *Running Lean: Iterate from Plan A to a Plan that works*. O'REILLY MEDIA, 2nd edition, 2012.
- [50] L. N. A. Media. Social media comparison infographic. <https://leveragenewagemedia.com/blog/social-media-infographic/>, 2014. [Online; sidst set d. 22-05-2014].
- [51] Netflix. Netflix - start your free month. <https://www.netflix.com/getstarted?locale=en-DK>, 2014. [Online; sidst set d. 06-06-2014].
- [52] F. A. Nielsen. Afinn: A new word list for sentiment analysis on twitter. <http://finnaarupnielsen.wordpress.com/2011/03/16/afinn-a-new-word-list-for-sentiment-analysis/comment-page-1/>, 2014. [Online; sidst set d. 16-06-2014].

- [53] F. A. Nielsen. distributed my 2477-word zip fil - afinn-111.txt. <http://www2.compute.dtu.dk/~faan/data/AFINN.zip>, 2014. [Online; sidst set d. 16-06-2014].
- [54] F. A. Nielsen. Github. https://github.com/abromberg/sentiment_analysis/tree/master/AFINN, 2014. [Online; sidst set d. 20-06-2014].
- [55] NLTK. Source code for nltk.classify.naivebayes. http://www.nltk.org/_modules/nltk/classify/naivebayes.html, 2014. [Online; sidst set d. 18-06-2014].
- [56] Overskrift.dk. Konkurrent forside. <http://www.overskrift.dk/>, 2014. [Online; sidst set d. 06-06-2014].
- [57] Pinterest. Pinterest grid. <http://www.streamcompanies.com/blog/advertising-topical/lets-get-gridiculous/attachment/pinterest-grid>, 2014. [Online; sidst set d. 01-06-2014].
- [58] B. N. på Twitter. Tv2breaking. <https://twitter.com/tv2breaking>, 2014. [Online; sidst set d. 22-05-2014].
- [59] Ritzau. Omdiskuteret valgvideo trukket tilbage. <http://www.dr.dk/Nyheder/Politik/EP14/Artikler/2014/05/13/095333.htm>, 2014. [Online; sidst set d. 22-06-2014].
- [60] Ritzau. Özlem cekic: Valgvideo skulle aldrig have været lagt op. <http://www.information.dk/497268>, 2014. [Online; sidst set d. 22-06-2014].
- [61] Salesforce. Benefits of saas. <http://www.salesforce.com/saas/benefits-of-saas/>, 2014. [Online; sidst set d. 11-06-2014].
- [62] Spamlaws.com. Types of spamming on twitter. <http://www.spamlaws.com/spamming-on-twitter.html>, 2014. [Online; sidst set d. 04-06-2014].
- [63] Spotify. Try premium for free. <https://www.spotify.com/dk/freetrial/>, 2014. [Online; sidst set d. 06-06-2014].
- [64] Stackoverflow. How many percent of the tweets does twitter sample api give? <http://stackoverflow.com/questions/13055370/how-many-percent-of-the-tweets-does-twitter-sample-api-give>, 2014. [Online; sidst set d. 04-06-2014].
- [65] T. L. Startup. Build, measure, learn. <http://theleanstartup.com/principles>, 2014. [Online; sidst set d. 27-06-2014].
- [66] TIME. How twitter turned out to be the only second screen we really need. <http://techland.time.com/2014/01/29/twitter-second-screen/>, 2014. [Online; sidst set d. 27-06-2014].

- [67] N. Y. Times. The plus in google plus? it's mostly for google. http://www.nytimes.com/2014/02/15/technology/the-plus-in-google-plus-its-mostly-for-google.html?_r=0, 2014. [Online; sidst set d. 22-05-2014].
- [68] N. Y. Times. The plus in google plus? it's mostly for google. <http://www.nytimes.com/2014/02/15/technology/the-plus-in-google-plus-its-mostly-for-google.html>, 2014. [Online; sidst set d. 07-06-2014].
- [69] Trak.io. Freemium vs. free: Why we ditched our free plan. <http://blog.trak.io/freemium-vs-free-why-we-ditched-our-free-plan/>, 2014. [Online; sidst set d. 07-06-2014].
- [70] Twitter. State of twitter spam. <https://blog.twitter.com/2010/state-twitter-spam>, 2010. [Online; sidst set d. 09-06-2014].
- [71] Twitter. Hashtag - drgrandprix. <http://u.uhrebrink.dk/11IWP7I>, 2014. [Online; sidst set d. 22-05-2014].
- [72] Twitter. Hashtag - ep14dk. <http://u.uhrebrink.dk/St0Fd0>, 2014. [Online; sidst set d. 22-05-2014].
- [73] Twitter. Start a conversation, explore your interests, and be in the know. <https://twitter.com>, 2014. [Online; sidst set d. 22-05-2014].
- [74] Twitter. Twitter fortæller hvordan man håndtere spam. <https://support.twitter.com/articles/79901-api-developers-abuse-prevention-and-security>, 2014. [Online; sidst set d. 22-05-2014].
- [75] Tænk. Apple trækker opladere retur. <http://taenk.dk/nyheder/iphone-ladere-traekkes-tilbage>, 2014. [Online; sidst set d. 23-06-2014].
- [76] T. UK. Total tweet for eurovision finals. <https://twitter.com/TwitterUK/status/465427409660289024>, 2014. [Online; sidst set d. 25-06-2014].
- [77] Unbounce. The benefits of using video on landing pages. <http://unbounce.com/landing-page-articles/the-benefits-of-using-video-on-landing-pages/>, 2014. [Online; sidst set d. 30-05-2014].
- [78] A. Universitet. Is the sample good enough? undersøgelse af kvaliteten på sample api. <http://bit.ly/1nbsBAQ>, 2014. [Online; sidst set d. 20-05-2014].

- [79] M. Vadskaer. Hashjacking - sjov med hashtags. <http://www.v4d5.net/blog/hashjacking>, 2014. [Online; sidst set d. 10-06-2014].
- [80] Version2. Twitter siger ja til patentdomstol. <http://www.version2.dk/artikel/twitter-siger-ja-til-patentdomstol-58767>, 2014. [Online; sidst set d. 22-05-2014].
- [81] Webopedia. Unstructured data. <http://u.uhrebrink.dk/1mCOHvN>, 2014. [Online; sidst set d. 27-06-2014].
- [82] S. Whitworth. Sentiment analysis in python using nltk. <http://www.sjwhitworth.com/sentiment-analysis-in-python-using-nltk/>, 2014. [Online; sidst set d. 18-06-2014].
- [83] Wikipedia. 5 whys. <http://u.uhrebrink.dk/SuQW69>, 2014. [Online; sidst set d. 10-06-2014].
- [84] Wikipedia. A/b test. <http://u.uhrebrink.dk/1lmiwPV>, 2014. [Online; sidst set d. 14-06-2014].
- [85] Wikipedia. Eurovision song contest 2014. http://en.wikipedia.org/wiki/Eurovision_Song_Contest_2014, 2014. [Online; sidst set d. 23-06-2014].
- [86] Wikipedia. Frequency distribution. http://en.wikipedia.org/wiki/Frequency_distribution, 2014. [Online; sidst set d. 04-06-2014].
- [87] Wikipedia. High- and low-level. <http://u.uhrebrink.dk/1mC5SYv>, 2014. [Online; sidst set d. 10-06-2014].
- [88] Wikipedia. Spamming. [http://en.wikipedia.org/wiki/Spam_\(electronic\)](http://en.wikipedia.org/wiki/Spam_(electronic)), 2014. [Online; sidst set d. 23-06-2014].
- [89] Xplenty. Konkurrent forside. <https://www.xplenty.com/>, 2014. [Online; sidst set d. 06-06-2014].

Bilag

A Google Cloud omkostninger



Figur 1: Google Cloud min. omk. Kilde <https://cloud.google.com/products/calculator/>

B TV2 Breaking News Tabel

Tweet	Delt	Favorit
Farlig brand i Ringsted: Politiet opfordrer beboere til at gå indenfor og lukke døre og vinduer.	4	0
Måling: Venstre nu blot tredjestørst til EU-valg #ep14dk	6	2
Københavns Vestegns Politi har torsdag anholdt og sigtet tidligere chefredaktør på Se og Hør, Henrik Qvortrup	9	3
Brøndby IF venter et underskud før skat på 160-170 millioner kroner, fremgår det af klubbens kvatalsregnskab	3	1
Ny måling: Venstre har ikke været mindre siden 2011 #dkpol	5	0
Lige nu: Andre EU-borgere skal ikke have fri adgang til danske velfærdsydelser, mener to ud af tre. #tv2valg	13	7
Lige nu: Ja til mere EU-politisamarbejde, også hvis retsforbeholdet lempes, siger hver anden i ny måling. #tv2valg	10	6
Fuld opbakning fra forretningsudvalg til Lars Løkke Rasmussen	1	2
Lige nu: 8 ud af 10 danskere ønsker kædeansvar for virksomheder. Flere EU-politikere er imod. Debat nu: #tv2valg	8	10
Dramatiske dage i #dkpol. Kl. 13 er det tid til spørgetime med statsministeren. #ftlive #tv2valg	1	0

C #twitterhjerne eksempler

Linse Daugaard @Linsedaugaard
Er der nogen der ved hvor man kan få slebet en køkkenkniv i dag? Aalborg og omegn? #twitterhjerne
21/06/14 09:08

Torben Tronborg @torbentronborg
Bedste svømmehal i Kbh. / storkbh. med aktiviteter, rutochabane, klatrevæg, leg og sjov til far/søn tur (5 år) ? #twitterhjerne
21/06/14 10:15

Tommy Bæk Seggaard @baeksoegaard
@Linsedaugaard Sving forbi @Kokkekniven :) 4d

Anders Lorentsen @anderslorentsen
@Linsedaugaard din lokale slagter måske? 4d

Kristian Petersen @Kokkekniven
@baeksoegaard @Linsedaugaard Ville gerne men er ude og grille til Mtb 12 timers løb i Skerping! Men kig gerne forbi til kaffe :) 4d

Linse Daugaard @Linsedaugaard
@anderslorentsen Har du prøvet det? 4d

Linse Daugaard @Linsedaugaard
@Kokkekniven @baeksoegaard Tak for tip og kaffetilbud :) 4d

Nepper @ahnepper
@torbentronborg @twhjerne Valby vandkultur... 4d

Lars Flindt Pedersen @flindstone
@torbentronborg @twhjerne har ikkeselv været der men ifølge flere pålidelige kilder er vandkulturhuset i Valby fedt. kulturogfritid.kk.dk/valby-vandkult... 3d

Torben Tronborg @torbentronborg
@ahnepper @flindstone Fedt, vi er på vej derud nu :) Tak for input ! pic.twitter.com/VNnkQ9sfaX 3d

Figur 2: To eksempler på #twitterhjerne

D Stopordsliste

span_words.txt					
1	androidgames	24	promocode	47	nowwatching
2	gameinsight	25	listeningto	48	selfie
3	ipadgames	26	bignaturaltits	49	crushproblemanight
4	TeamFollowBack	27	bigboobs	50	candycrush
5	TFBJP	28	bignaturals	51	workswag
6	openfollow	29	bigmelons	52	swag
7	mustfollow	30	porn	53	curves
8	50ADay	31	hardcore	54	curvy
9	THF	32	juggs	55	chick
10	HITFOLLOWSTEAM	33	busty	56	phat
11	NowPlaying	34	hugeboobs	57	freebiefriday
12	35giveadays	35	bbw	58	teamfatbaby
13	Giveaway	36	voluptuous	59	cum
14	striptease	37	bitcoin	60	cumfacial
15	whooty	38	rtx	61	sex
16	fetish	39	gay	62	follow
17	bigtits	40	lesbian	63	harvested
18	twitition	41	amreading	64	collected
19	instaphoto	42	carsale	65	coins
20	instalive	43	welcometweet	66	iphone
21	instalove	44	2nelcreush	67	retweet
22	followback	45	followfriday		
23	dating	46	pussy		

Figur 3: Stopordsliste

Antal	Ord	Antal	Ord
708	follow	115	apple
510	rt	73	my
354	Gain	70	iphone
181	ipad	38	ipad
179	me	37	me
179	retweet	36	phone
178	back	28	GuitarmanDan
177	MUST	27	half
177	everyone	27	old
94	Love	26	hour
62	ANDROID	26	play
32	BELIEBER	26	year
26	FOLLOWBACK	25	Disneys
22	ONEDIRECTION	25	Frozen

Tabel 1: Frekvensfordeling inden og efter fresortering af hvid støj

	Apple		Amazon		ATT		Cloudmagic		Digitalocean
130417	iphone	47140	via	24262	iphone	42	email	260	deploy
41003	ipad	17152	reply	22122	new	39	app	259	enter
23701	like	12428	free	20321	battery	20	android	259	free
18714	one	9066	sample	9927	amazon	20	great	256	ssd
18707	im	7833	phone	9909	apple	16	iphone	255	vps
17655	using	7167	amp	9509	heart	14	best	249	promo
16733	juice	6081	bought	8992	matthew	13	client	245	credit
16733	phone	6003	add	8770	release	12	thanks	244	code
16385	new	5714	book	8742	crazy	11	guys	244	time
16270	get	5401	get	8690	schultz	11	mail	235	gt
13191	got	5115	new	7099	samsung	9	access	226	new
12592	store	5095	fire	7071	look	9	emails	193	ideas
11759	amp	4271	like	6555	phone	9	work	177	dodroplet
11668	pie	3977	buy	6375	case	8	cloud	122	http
11407	tao	3729	x	6342	nice	8	gmail	107	expires
11259	app	3712	listed	6321	mad	8	good	82	data
10875	want	3506	one	6309	meet	7	amp	82	server
10825	dont	3328	im	6200	london	7	hello	73	like
10489	android	3072	love	6190	streets	7	ios	71	thanks
10166	case	2874	check	6179	hatter	7	like	67	stack

	Eurovision		McDonalds		Nokia		Samsung
988	song	23556	breakfast	3148	never	4313	galaxy
727	like	20537	know	2984	mess	2360	phone
559	poland	20360	day	2943	win	1748	iphone
553	conchita	20203	let	2106	chance	1392	note
547	year	19326	walk	1567	phone	1381	one
490	im	18999	around	1315	amp	1222	new
480	won	18563	folk	1231	lumia	1163	phones
462	winner	18205	bitches	1100	support	1016	charger
437	butter	16571	want	772	bag	964	like
426	churn	9239	im	772	exclusive	936	usa
401	next	8518	like	738	accessories	884	get
400	love	8237	get	737	filled	880	im
377	read	7822	food	734	mystery	853	device
361	contest	7018	amp	692	trousers	823	amp
360	uk	6722	mom	599	new	741	please
337	get	5764	go	557	x	731	got
332	win	5694	home	521	cmon	713	news
322	dont	5692	yo	516	boys	684	apple
301	good	5037	stop	482	portable	667	case
299	songs	4976	say	479	like	585	good

Figur 4: Top 20 lister for mest brugte ord på brands

E Iteration fire - Wordclouds



Figur 5: Top 20 Wordcloud for Amazon



Figur 6: Top 20 Wordcloud for Apple



Figur 7: Top 20 Wordcloud for ATT



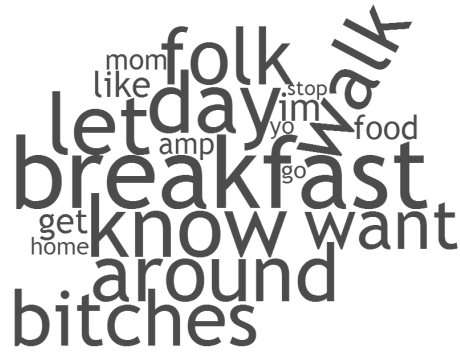
Figur 8: Top 20 Wordcloud for Cloud Magic



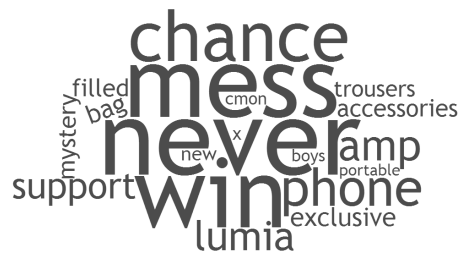
Figur 9: Top 20 Wordcloud for Digital Ocean



Figur 10: Top 20 Wordcloud for Eurovision



Figur 11: Top 20 Wordcloud for MCDonalds

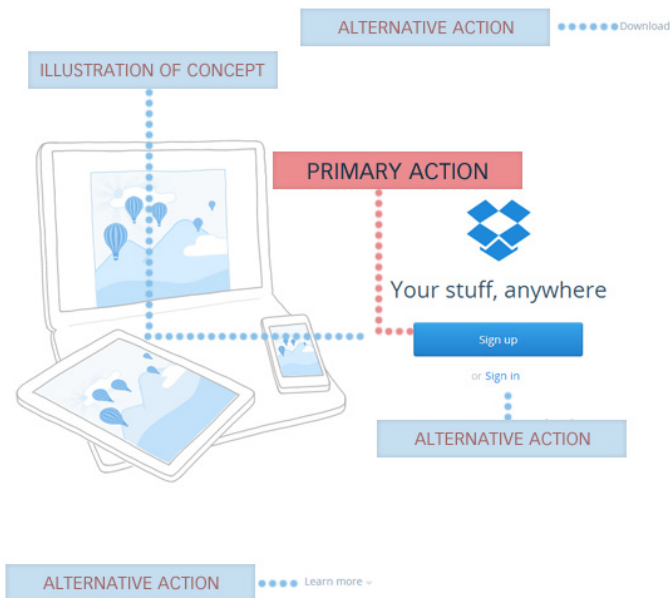


Figur 12: Top 20 Wordcloud for Nokia



Figur 13: Top 20 Wordcloud for Samsung

F Landing page cases

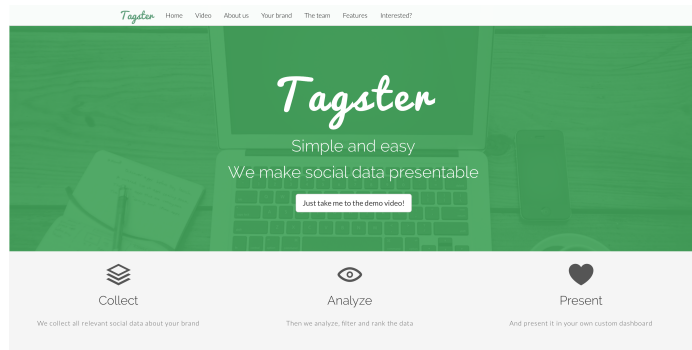


Figur 14: Landing Page Case: DropBox. Kilde <http://my.umbc.edu/groups/web-dev/news/36650>

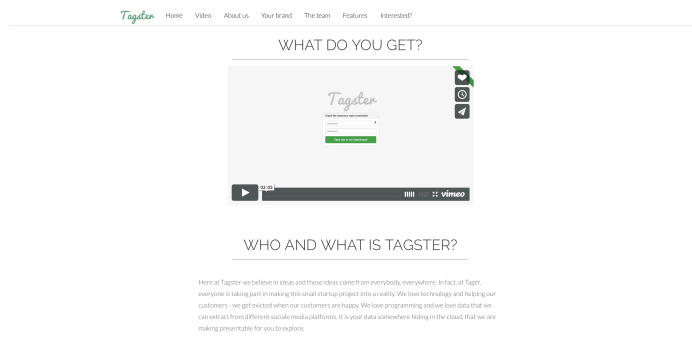


Figur 15: Landing Page Case: Mailbox. Kilde <http://my.umbc.edu/groups/web-dev/news/36650>

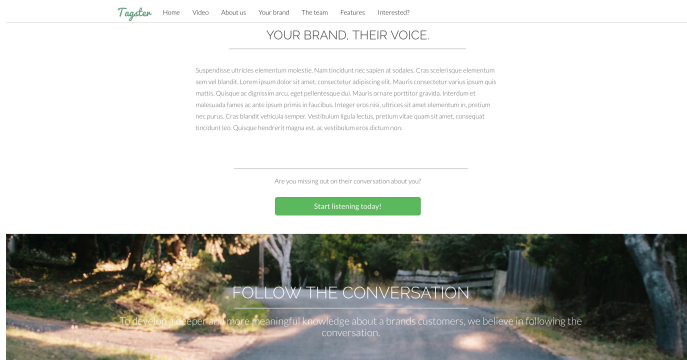
G Landing page - Tagster.dk



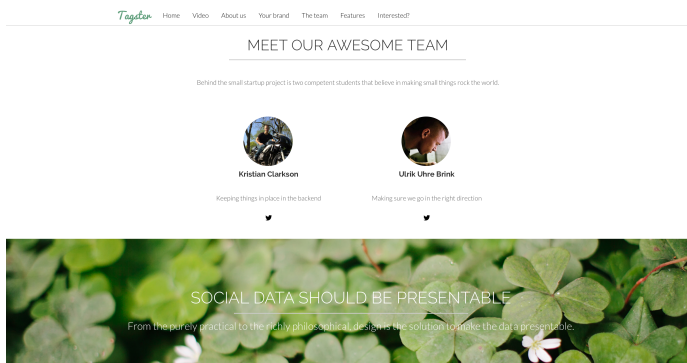
Figur 16: Tagster.dk - Header



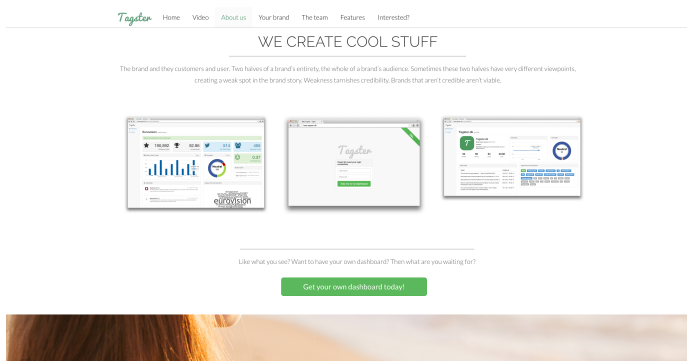
Figur 17: Tagster.dk - Video gennemgang



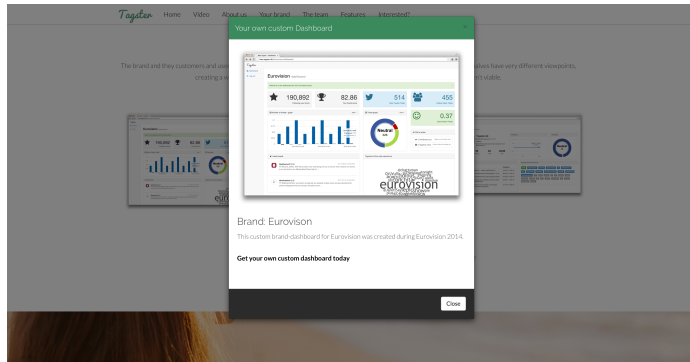
Figur 18: Tagster.dk - Dit brand, deres stemme



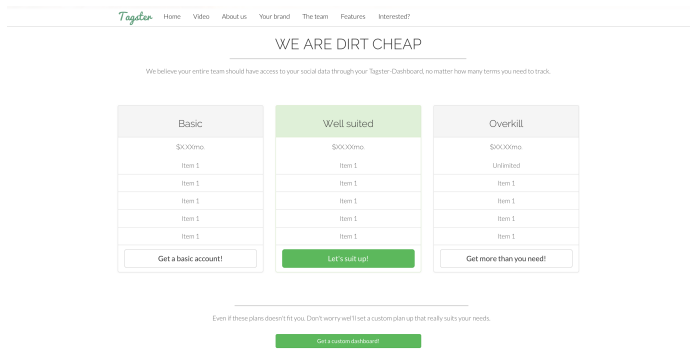
Figur 19: Tagster.dk - Hvem er vi?



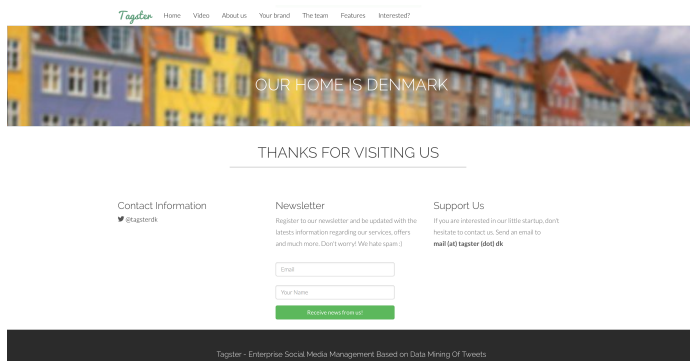
Figur 20: Tagster.dk - Features



Figur 21: Tagster.dk - Features - Dashboard



Figur 22: Tagster.dk - Prissætning



Figur 23: Tagster.dk - Kontakt

H Landing page - Resultater

The screenshot displays a Twitter thread with the following content:

- Ulrik Uhre Brink** (@UlrikUhreBrink) 01/04/14: Dagens #spyslot af @tagsterdk pic.twitter.com/7SuByeGn3h
- Lasse Kjær C.** (@lassekjaertw) 01/04/14: @UlrikUhreBrink @tagsterdk Glæder mig vildt meget til at se mere.
- Tagster.dk** (@tagsterdk) 01/04/14 14:57: @lassekjaertw @ulrikuhrebrink og vi glæder os helt vildt til at vise dig vores løsning.
- Lasse Kjær C.** (@lassekjaertw) 01/04/14: @tagsterdk @UlrikUhreBrink Kan I sige noget om, hvornår man kan se mere?
- Tagster.dk** (@tagsterdk) 01/04/14: @lassekjaertw @ulrikuhrebrink vi har ikke en ETA på projektet endnu. Du kan jo være med i test-fasen og få en gratis konto #giveaway
- Lasse Kjær C.** (@lassekjaertw) 01/04/14: @tagsterdk Gerne! Det kunne fungere rigtig godt. Har efterhånden indsigt i mange alternative løsninger, så vil gerne være med i beta.
- Tagster.dk** (@tagsterdk) 01/04/14: @lassekjaertw så snart vi har en working beta, så får du login.
- Lasse Kjær C.** (@lassekjaertw) 01/04/14: @tagsterdk Lyder spændende. Det vil jeg glæde mig til. :)
- Tagster.dk** (@tagsterdk) 01/04/14: @lassekjaertw jo mere input vi kan få fra jer brugere, desto bedre og mere præcis bliver vores løsning.
- Lasse Kjær C.** (@lassekjaertw) 01/04/14: @tagsterdk Dejligt at høre I også tror på den filosofi. :)

Figur 24: Første interessede følger



Figur 25: A/B splittest feedback

Tagster Home Video About us Your brand The team Features Interested?

SOCIAL DATA HAS NO BOUNDARIES

Social data is like the Wild West and unexplored but rife with opportunity. We filter and extract your relevant data for you to explore.

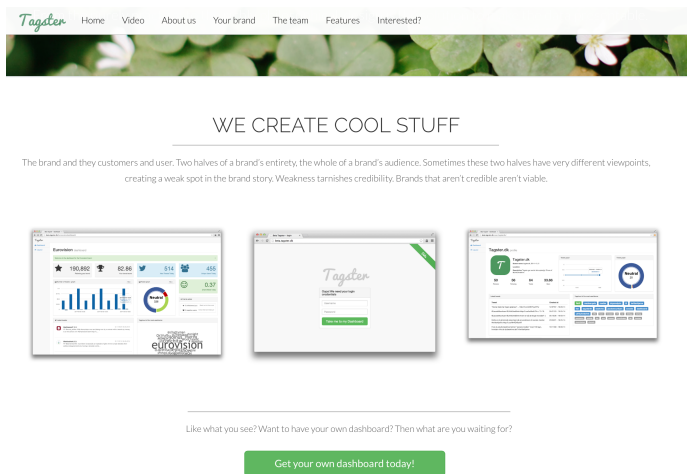
YOUR BRAND. THEIR VOICE.

Figuring out your brand voice for social media can be a challenge if you're just starting out, but that doesn't mean you should let analysis paralysis stop you from rocking it on social networks. Your customers are your key to success - happy customers and soon-to-be customers are your key to success. It is your brand, but it is their voice that set the standards. Do you listen to what they are saying about your brand? If not, you are really missing out on valuable information.

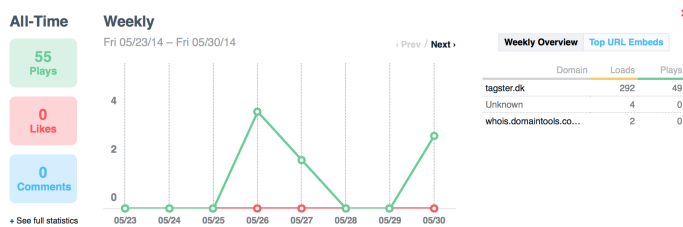
Are you missing out on their conversation about you?

Start listening today!

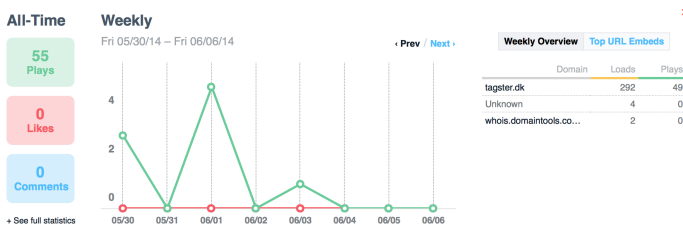
Figur 26: Your brand. Their voice. Start listening today!



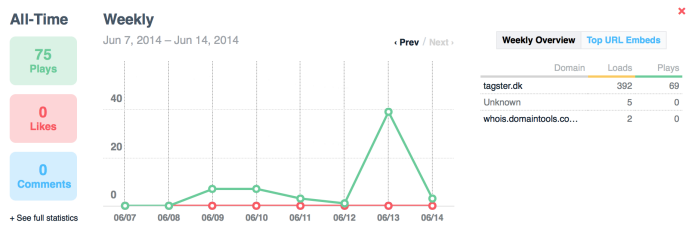
Figur 27: We create cool stuff. Get your own dashboard today!



Figur 28: Tagster demo-video resultater for uge 1

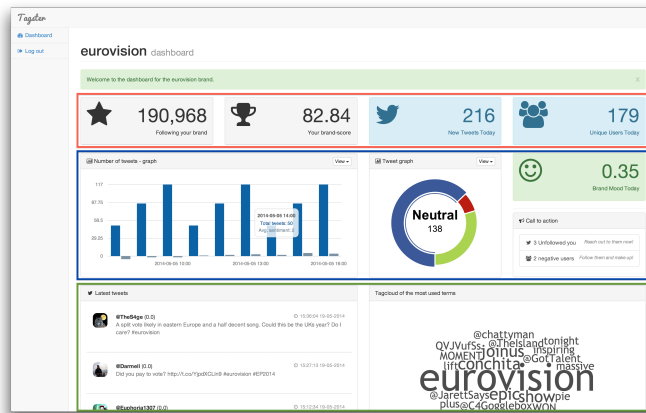


Figur 29: Tagster demo-video resultater for uge 2

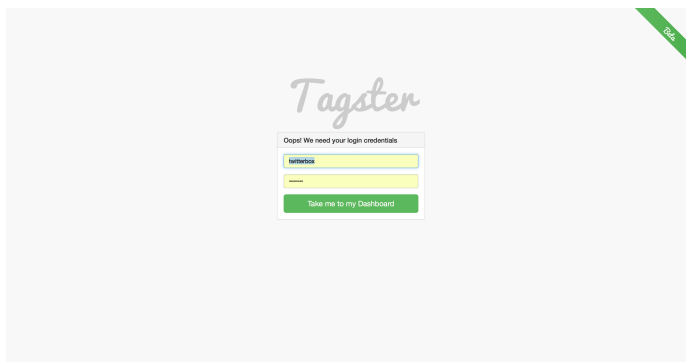


Figur 30: Tagster demo-video resultater for uge 3

I Web-applikation - Tagster.dk



Figur 31: Tagster.dk - Opdeling af dashboard



Figur 32: Tagster.dk - Log ind på dit dashboard

The screenshot displays a grid of tweets from the user 'voteman' on the Tagster.dk platform. The interface includes a sidebar with navigation options like 'Dashboard' and 'Log out'. The main content area is titled 'Tweets voteman' and shows a list of tweets. Each tweet is presented in a card format with a header containing the user's name and the time of the tweet. The tweet text is followed by a link to the content. Below each tweet, there is a summary of user interactions, such as 'The tweet is positive', 'The tweet is neutral', or 'The tweet is negative', along with counts for likes, retweets, and replies. The tweets themselves contain various links to external content, including YouTube videos, political articles, and social media posts. The overall layout is clean and organized, typical of a social media feed.

Figur 33: Tagster.dk - Udforsk - Voteman

J Web-applikation - betatest

Betatest af Tagster

Hvad bruger vi nu?

Til at monitorere hashtags benytter vi, på nuværende tidspunkt, Topsy til dette formål. Topsy giver et hurtigt sammenligningsgrundlag mellem hashtags og kan, med deres minut-for-minut-feature, give et detaljeret indblik i Twitter data ved store begivenheder. Topsis overskuelighed gør det nemt, at generere et visuelt overblik/sammenligning mellem 2-4-5-eller flere hashtags.

Vi har, ved tidligere lejligheder (blandt andet i DR) også benyttet Falcon Social. Falcon Social er et mere komplekst system der favner bredere, på tværs af sociale netværk. Dets features er bygget op i Listening og Analytics der, i vores optik, er de mest interessante. Analytics har en 'rapportbygger' der kan udarbejde et grafisk overblik over udvalgte Twitter/FB/YouTube/Instagram sider. Analytics fungerer godt til Twitter, da data fra Twitter selv er begrænsede. Det samme med Instagram. Men for eksempel til FB rapporter er Facebooks egen Insights at foretrække. YouTubes analyseværktøjer har vi ikke arbejdet så meget med.

Præsentation af data

Dashboardets fire øverste felter giver god mening ifht. overblik. Fin måde at præsentere 'overskrifter' på. 'Brand Score' får mig til at tænke på Falcon Socials 'Engagement Level' der, i min optik, ikke giver den store mening. Falcons Engagement Level har en kompliceret formel for udregning. Hvad er Tagsters bagvedliggende udregning for brand-score? Jeg kan ikke trykke på det, jeg kan ikke se flere detaljer. Jeg synes den er overflødig.

I øvrigt er Engagement Level (eller brand-score) umiddelbart ikke noget, jeg ville anvende i en sammenligningssituation. Her ville jeg hellere kigge på reel engagement (retweets, mentions, favorites, follows) end en udregnet score. Brand-mood er lidt det samme. 0.17 i brand-mood – hvad betyder det? Hvad betyder tallet? Farven, og smileyen, giver mening.

Number of tweets mangler muligheden for at gå tilbage i tiden. Jeg kunne godt tænke mig, at kunne gå en måned tilbage i menuen. Custom range, a'la Topsy, kunne være helt, helt perfekt. Det er virkelig en god feature og anvendelig feature. Det gør det for eksempel muligt at sammenligne sidste års Eurovision, DMGP, X Factor eller sådan.

Call to action er en helt vildt god ide. Hurtigt, nemt og meget sådan to-do-agtig. I like. Det samme med Tagcloud'en. Det er ord, eller brugernavne, indenfor #Eurovision hashtagget – ja? Det kunne måske være interessant at kunne vælge enten ord, eller hashtags der også bliver brugt sammen med #Eurovision.

Rød, hvid og grøn farve på tweets er en god ide til at signalere stemning. Hvordan afgøres farven? Interessant. Men rigtig fin metode til at give overblik. Ville også give god mening i en social tv-sammenhæng hvor der skal findes tweets til skærm.

User

Lidt små UI problemer ved profilvisning, men klart. Det er beta.

Jeg synes der skal være en grafisk forskel på "followers, following, tweets og score". Tagcloud'en kunne godt være opdelt i hashtags og reelle ord. Umiddelbart er jeg mest tilhænger af denne grafiske fremstilling af tags/ord fremfor selve 'cloud-grafikken' der findes i Dashboard.

Tweets

Umiddelbart den, af de tre, sider der giver mindst overblik. Der er meget tekst, udover selve tweetet, der virker forstyrrende for øjnene. Farvekoderne (grå, grøn, rød) indikerer værdien af tweetet, hvorfor 'The tweet is neutral' eller 'The tweet is positive', i min optik, er overflødig. Rank behøver ikke at have supplerende tekst, blot vise værdi (Klout). Det kunne være brugbart, hvis der blev highlightet alt over Klout 60 for eksempel. Så vidste man, at der var noget man skulle være opmærksom på.

Pris

Som udgangspunkt vil jeg mene at din basispris på 499 er for høj, hvis målet er små virksomheder, der ikke er sikre på, om det er en idé. Tilbyd evt. en 7, 14 eller 30-dages gratisperiode. Hos 140tegn.dk ville vi eks. ikke være parat til at kaste 499 efter noget, vi ikke havde testet og nok heller ikke med de funktioner det har.

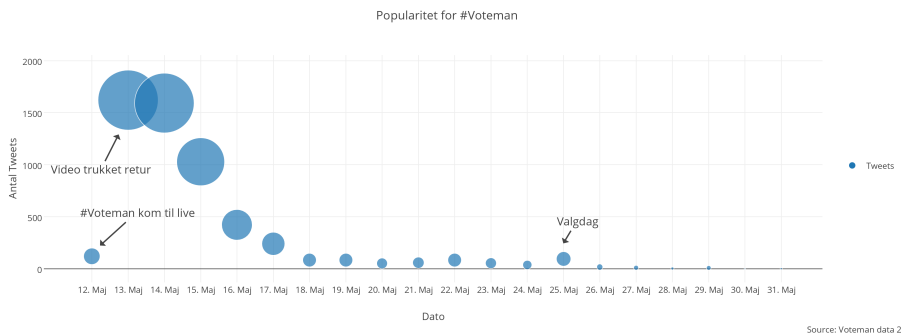
Springet op til 999 er meget småt. Er 499-999 virkelig en barriere? Virksomheder, som har råd til 500/md har også råd til 1000. På den måde kan springet eks. sagtens være til 1500 allerede her. De mest prisfølsomme kunder, er de, der ville købe den billigste. Altså bliver den prismæssige barriere lavere og lavere i takt med du kommer op i pris. Mellem 1-100 kr. er elasticiteten høj (du kan aflæse totalsalg direkte af bare en kroners forskel). Men mellem 500-1.000 kr. er det markant sværere at afgøre hvad noget er værd. Altså vil de kunder, der køber noget til 500 kr. også være indstillet på at købe noget til 1.000 kr.

Derfor ville jeg sætte basispakken ned eks. til 14 dages gratis og derefter 199 kr./md. (måske endda ENDNU mere skrabet i funktioner).

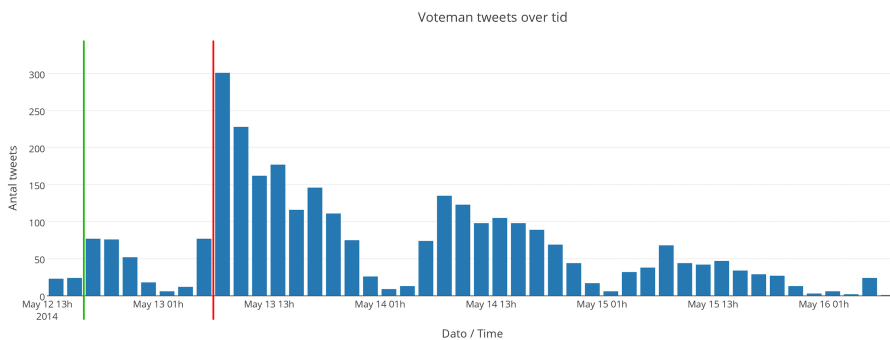
Mellempakken kan sagtens sættes højere. Eks. til 1499.

Hvad angår dit "prima-produkt". Virksomheder som DR eller store medier/bureauer er vant til at betale markant mere for software, der virkelig giver en høj værdi. Det er svært at sige helt præcis, men min erfaring siger mig, at du roligt kan prissætte 5000+ eller måske endda slet ikke prissætte og skrive: "Skræddersyet" og forhandle prisen med den type kunder individuelt.

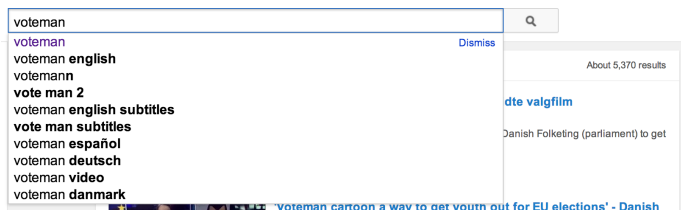
K Case - Voteman



Figur 34: #Voteman popularitet før, under og efter valget



Figur 35: #Voteman over tid



Figur 36: #Voteman på Youtube

Dato	Antal tweets
12. Maj	121
13. Maj	1622
14. Maj	1593
15. Maj	1029
16. Maj	423
17. Maj	241
18. Maj	84
19. Maj	84
20. Maj	52
21. Maj	59
22. Maj	84
23. Maj	54
24. Maj	38
25. Maj	95
26. Maj	17
27. Maj	11
28. Maj	4
29. Maj	10
30. Maj	1
31. Maj	2
1. Juni	2
2. Juni	3
3. Juni	4
4. Juni	1
5. Juni	4
6. Juni	2
7. Juni	5
8. Juni	1
9-14. Juni	7
15. Juni	2
16. Juni	3
17. Juni	1
18. Juni	3
19. Juni	2
20. Juni	2
22. Juni	5
Total	5672

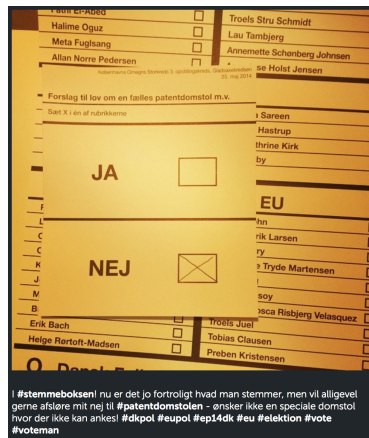
Tabel 2: Antal tweets for #voteman pr. dag

L Case - Voteman - Valgfie



Tak til #voteman

Figur 37: #Voteman #Valgfie



#stemmeboksen! nu er det jo fortroligt hvad man stemmer, men vil alligevel gerne afsløre mit nej til #patentdomstolen - ønsker ikke en speciale domstol hvor der ikke kan ankes! #dkpol #eupol #ep14dk #eu #elektion #vote #voteman

Figur 38: #Voteman #Valgfie

M Case - Voteman - Instagram/Twitter



Puk Falkenberg
@PuuksF

@UlrikUhreBrink Jeg stemmer uanset hvad, men synes klart det var en go ny måde at fange folk på. :D



Helene Kristensen
@Heleneken

@UlrikUhreBrink Nej. Mine forældre og min egen samvittighed fik mig til at stemme.



Philip Christensen
@Philipo

@UlrikUhreBrink Nej, for vidste jeg ville stemme inden jeg hørte om Voteman.



Tobias Hydeborg
@TobiasHydeborg

@ulrikuhrebrink Nope. :-)



Puk Falkenberg
@PuuksF

@UlrikUhreBrink > Jeg grinte og synes det var hyle morsomt de havde lavet den. Elsker skæve indslag der får mig til at grine :)



Helene Kristensen
@Heleneken

@UlrikUhreBrink At den slet ikke er henvendt til mig. Det er slet ikke min type humor.



Philip Christensen
@Philipo

@UlrikUhreBrink Jamen jeg kunne godt lide den. Sjov og fart over feltet :)

Figur 39: #Voteman holdninger på Twitter



Figur 40: #Voteman's indvirkning på Instagram