

Når ordvalgene avslører skjulte hensikter og markedstrender

Stadig flere analytikere og aksjefond forsøker å skaffe seg en fordel gjennom å analysere de enorme mengdene ord som individer og selskaper etterlater seg på internett, i mediene og i tradisjonell selskapskommunikasjon.



Assistant Professor of Accounting
Kyrre Kjellevold
Norges Handelshøyskole

Om Forskning og revisjon

I denne faste spalten vil Assistant Professor Kyrre Kjellevold ved Norges Handelshøyskole drøfte nyere forskning i finans, regnskap og revisjon, og hvordan det kan ha påvirkning på revisors arbeid.

I dette innlegget vil jeg drøfte noen av de mulighetene og utfordringene man står overfor når man tar i bruk moderne tekstanalyseverktøy.

Gammel metode i ny drakt

Tekstanalyse er ingen ny oppfinnelse. Siden 1300-tallet har både munkene og andre lærde søkt å oppdage skjulte sammenhenger ved å studere frekvensen og bruken av spesifikke ord i alt fra Bibelen og Shakespeares verker til politiske taler.

I dag har tilgangen til enorm datakraft og det store antallet tilgjengelige informasjonskilder gitt tekstanalyse en ny renessanse. I stedet for munkene som skummer gjennom religiøse tekster i dunkel belysning, kan en algoritme skumme gjennom millioner av tekstdokumenter på få minutter.

Nyere tekstanalyseteknikker kan for eksempel:

- Måle frekvensen av ord som er negativt eller positivt ladede, og dermed gi en indikasjon på om rapporten f.eks. inneholder dårlige nyheter.

- Benytte maskinlæringsteknikker til å undersøke om spesifikke ordvalg fører til endringer i investorers oppfatning om markedet og selskapet.
- Analysere om ledere ofte benytter spesifikke setninger eller ord når de ønsker å tilsløre tap eller på andre måter manipulere regnskapet.

Positivt eller negativt ladet tekst?

Den første bruken av tekstanalyse i finans- og regnskapsforskning gikk ut på å kartlegge hvilke ord som har en negativ og positiv betydning i finansiell kommunikasjon. Frekvensen av slike ord ble benyttet til å måle hvorvidt en årsrapport var negativt eller positivt ladet.

De negative ordene har gjerne mest betydning, fordi mennesker ofte har en generell tapsaversjon og siden de positive ordene ofte blir brukt til å tilsløre dårlige nyheter. Positive ords betydning kan derfor være mer uklar og avhengig av hvilket medium man analyserer.

Slik tekstanalyse vanskeliggjøres også av det som er kjent som Zipfs lov: Et lite antall ord vil dominere når du tel-

ler hvor ofte ord forekommer i en tekst. Da må man være påpasselig når man fører en algoritme med informasjon om hva et ord betyr, siden det kan få store utslag dersom et ord forstås feil. Standard ordlister utviklet av språkforskere kan ofte ikke benyttes siden de ikke er tilpasset finansiell kommunikasjon. Eksempelvis er ordet skatt ofte negativt for en privatperson, men dagligdags for et selskap.

Når man tar hensyn til slike potensielle fallgruver, finner man i analyser av språkbruken i årsrapporter og/eller i investordiskusjoner på Twitter og andre sosiale medier, at negativt ladet stemming samvarierer med kursfall, konkurrisiko og andre uheldige konsekvenser for et selskap.¹ Tekstanalyser kan hjelpe oss å forstå hvor man tror markedet skal gå, og gi oss mer informasjon enn å lese en analytikerrapport.

Maskinlæring og finansnyheter

En gruppe amerikanske forskere har undersøkt hvorvidt man basert på maskinlæring kan trene algoritmer til å

¹ De oftest forekommende negative ordene i amerikanske 10-Ks er: *loss, losses, claims, impairment, against, og adverse*. Les Loughran og McDonald (2011).

ABONNEMENT 2021



REVISJON OG REGNSKAP

Norges ledende fagblad innen revisjon, regnskap, selskapsrett, skatt og avgift. Målgruppe: Revisorer, regnskapsførere, jurister og andre faginteresserte. Abonnementet gir tilgang til nettutgaven av bladet. Arkiv tilbake til nr. 5/2003. 8 utgaver pr. år.

Priser årsabonnement

Ordinær kr 860,-
Student kr 400,-



REVISOR INFORMERER

Et informativt og lettlest temablade innen skatt, avgift, selskapsrett, regnskap, arbeidsmiljø og personlig økonomi. Målgruppe: Eiere og ledere i små og mellomstore bedrifter. 4 utgaver pr. år.

Priser årsabonnement*

Medlem DnR kr 200,-
Ikke-medlem kr 260,-
Student kr 95,-



FACTA ÅRSOPPGJØRET

Omfatter de viktigste regnskaps- og skatteregler som behøves i forbindelse med årsoppgjøret. Målgruppe: Revisorer, ledere og ansatte i økonomi- og regnskapsavdelinger.

Priser årsabonnement*

Medlem DnR kr 200,-
Ikke-medlem kr 280,-
Student kr 95,-

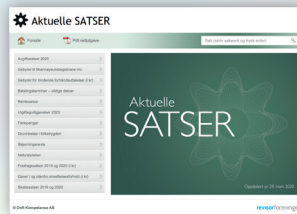


KOMBIABONNEMENT

Består av Facta Årsoppgjøret og fire utgaver av Revisor informerer. Målgruppe: Eiere, ledere og ansatte i bedriftens økonomi- og regnskapsfunksjon.

Priser årsabonnement*

Medlem DnR kr 340,-
Ikke-medlem kr 450,-
Student kr 145,-



AKTUELLE SATSER NETTUTGAVE

Webbløsning med oppdaterte skatte- og avgiftssatser, betalingstermin, bilgodtgjørelse, naturallytelse mv.

Pris årsabonnement

Medlem DnR kr 1500,-



eFACTA

Med eFacta kan du gi alle dine kunder tilgang til reglene som gjelder for årsoppgjøret 2020.

Pris årsabonnement

Medlem DnR kr 4000,-



NYTT FRA REVISOR

Gi kundene dine et nyhetsbrev! Nå kan du bestille abonnement på nyhetsbrevet Nytt fra revisor som gir dine kunder løpende oppdateringer gjennom året!

Pris årsabonnement

Medlem DnR kr 3000,-

* Pris eks. porto.

benytte språket i finansnyheter til å predikere endringer i aksjemarkedet.

Deres algoritme tråler gjennom millioner av historiske finansnyheter og trenes på sammenhengen mellom teksten i nyhetsartikkelen og kurssvingningene til selskapet som omtales. De finner at en rekke artikkelord har negativ betydning for børskursen, der de med størst prediksjonskraft er: *shortfall, downgrade, disappointing, auditor, tumble, blame, hurt, plunge, og slowdown*. Det er altså klart uheldig for et selskap at revisor nevnes i samme artikkel.

Basert på algoritmen lager forskerne en porteføljestrategi der de kjøper de 50 selskapene som har høyest andel ord som rangeres som de mest positive, og shorter de 50 selskapene med høyest andel ord som rangeres som de mest negative. Strategien gir en realisert Sharpe Ratio på 4,29, og forskerne viser dermed at de tekstanalytiske prediksjonene også er økonomisk signifikante.

Uærlige selskapsledere og strategiske ordvalg

For revisorer er kanskje det aller viktigste hvorvidt bruk av tekstanalyseverktøy kan lede til bedre risikovurderinger og bidra til at man avdekker flere feil i regnskapet. Selv om forskningen er i en tidlig fase, viser studier at det å analysere tekst kan gi et verdifullt bidrag til prediksjoner av regnskapsmanipulasjon.

For eksempel velger ofte amerikanske ledere som vi vet (basert på etterpåklokskap) har manipulert regnskapene å distansere seg fra årsrapporten ved at deres navn sjeldnere forekommer i diskusjonen av strategier og fremtidige planer. Det kan være en strategi for å tilsløre eventuelle skyldspørsmål i etterkant.

Videre viser forskning at selskapsledere som ønsker å villede investorer og analytikere på kvartalstallspresentasjoner har en tendens til å benytte flere svært positivt ladede ord, færre referanser til aksjønærverdier og flere referanser til generell (og ikke bransjespesifikk) kunnskap. Markedet reagerer sterkere når slike ledere tar store avsetninger i regnskapet,

som gir en solid pekepinn på at investorer forstår at noen forsøker å villed dem.

Språklige barrierer og endret selskapskommunikasjon

Uten tvil inneholder årsrapportene i dag stadig mer tekst, samtidig som de trolig ikke blir lest så ofte som man skulle tro.² Overordnet reagerer også markedet lite når årsregnskap slippes både i USA og Norge, selv om endringer i regnskapet har stor prediktiv kraft i markedet.³ Ligger årsaken til at investorer er trege til å ta til seg ny verdifull informasjon i at årsrapportene har blitt noen ugjennomtrengelige mursteiner?

Lesbarhet er vanskelig å måle, spesielt fordi de målevariablene vi har fra annen språkforskning ikke tar hensyn til at årsrapporter inneholder mange komplekse ord som er godt kjent i markedet og som i stor grad avhenger av hvilken bransje selskapet opererer i.

Amerikanske selskapsregnskaper (10-Ks) har blitt fem ganger så store mellom 2005 og 2017, og språklige endringer i årsrapportene skjer nå 12 ganger så ofte som i 2005. Nyere analyser peker på at tre temaer utgjør en stor del av den økte tekstlengden: beskrivelser av internkontroll, risikostyring og regnskapsestimater.⁴ Faktisk viser det seg at et så enkelt mål som antallet megabytes en årsrapport utgjør, kan være et godt mål på lesbarheten, og at antall megabytes samvarierer med volatiliteten til selskapets aksje og analytikerens prediksjonsfeil. Kanskje viser det at man sliter med å tolke informasjonen.

Selskaper tilpasser seg robotene

Til slutt kan en spørre seg om språkbruken vil endres i tråd med at algoritmer utgjør større deler av lezerskaren.

2 Faktisk viser en studie av Loughran & McDonald (2017 – *The Use of EDGAR Filings by Investors*) at selskapsregnskapene (10-Ks) til amerikanske børsnoterte selskaper i snitt blir lastet ned 28 ganger i perioden rett etter at det blir sluppet på EDGAR (den amerikanske versjonen av Brønnøysundsregistrene). Ikke mye for et land med verdens største og mest aktive aksjemarked og over 300 millioner innbyggere.

3 Sjekk ut Cohen, Malloy, and Nguyen (2020) – *Lazy Prices*.

4 Les Dyer, Lang, and Stice-Lawrence (2017) – *The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation*.



Selv om forskningen er i en tidlig fase, viser studier at det å analysere tekst kan gi et verdifullt bidrag til prediksjoner av regnskapsmanipulasjon.

Begynnende forskning finner at amerikanske selskaper som ser at algoritmer utgjør en betydelig del av de som laster ned årsrapporten, tilpasser språket i årsrapporten ved f.eks. å redusere antallet negative ord og tilpasse toneleiet og ordbruken på selskapspresentasjoner.⁵ Ledere som vet at maskiner lytter, vil kunne ha interesse av å fremstå mer entusiastiske og positive på en balansert måte.

Språkbruk har alltid vært viktig, men i fremtiden vil selskaper kanskje lytte mer til hva robotene er opptatt av enn dyre PR-rådgivere. Utvilsomt vil selskaper finpusse ordvalgene mer i årene fremover.

Det er uansett en spennende tid for alle som jakter på verdifull informasjon som kan gi dem et lite forsprang i markedet.

Les videre (som en start):
Loughran og McDonald (2011) – *When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks*.

Loughran og McDonald (2016) – *Textual Analysis in Accounting and Finance: A Survey*.

Ke, Kelly og Xiu (2019) – *Predicting Returns with Text Data*.

Cao, Jiang, Yang og Zhang (2020) – *How to Talk When a Machine is Listening: Corporate Disclosure in the Age of AI*.

5 Dagens algoritmer kan også analysere muntlig språk, og kan dermed handle basert på analyser av opptak fra eksempelvis presentasjoner av kvartalstall.