

EXAMENSARBETE Forecasting Financial Indices from Financial News**STUDENT** Gustaf Backman**HANDLEDARE** Pierre Nugues (LTH), Edvard Sjögren (Kidbrooke Advisory)**EXAMINATOR** Marcus Klang (LTH)

Slå börserna med morgontidningen?

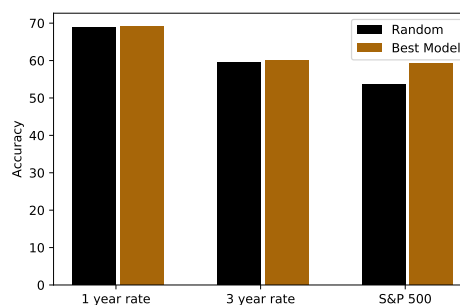
POPULÄRVETENSKAPLIG SAMMANFATTNING Gustaf Backman

Snabb och träffsäker hantering av data är av stort intresse inom många områden idag. De senaste åren har särskilt metoder inom automatisk språkhantering sett stora framsteg. I detta arbete visar jag hur deep learning och nyhetstitlar kan användas för att förutspå rörelsen på börserna med en träffsäkerhet på 59 %.

För att förstå hur en finansiell tillgång utvecklar sig över tid använder idag företag ofta tidsseriemodeller som bara tar hänsyn till tidigare värden av exempelvis en aktie eller ränta. Dessutom sitter någon och läser vad som händer i världen och tolkar hur det påverkar finansiella index. Detta tar tid och känns en aning ineffektivt, kanske kan det automatiseras och förbättras?

Mitt examensarbete provar olika sätt att tolka rubriker från finansiella nyheter i USA och undersöker hur de påverkar det amerikanska börsindeket S&P 500 samt två obligationsräntor. Jag har använt olika modeller för att göra om text till siffror. Både äldre varianter som i princip bara räknar vilka ord som finns i rubriker från en dag, samt modernare modeller såsom BERT, utvecklad av Google 2018. Dessa modeller för texttolkning kombineras sedan med någon typ av matematisk modell för att avgöra om ett index går upp eller ner, exempelvis logistisk regression.

Mina resultat visar att det går att förutspå om börsindeket S&P 500 går upp eller ner följande dag, åtminstone jämfört med att låta en bläckfisk gissa som i fotbolls-VM 2010. Vad gäller de två obligationsräntorna kan man dock lika gärna singla slant enligt mina resultat.



Träffsäkerhet på om ett index går upp eller ner kommande dag. Min modell är inte bättre än slumpen för de två räntorna, men bättre för S&P.

Den bästa modellen för S&P 500 var TF-IDF, en sorts normaliserad variant av ordräkning, tillsammans med en random forest-klassificerare. Båda dessa modeller har ganska många år på nacken, men verkar fungera bra till just denna applikation.

Dessa metoder kan exempelvis användas för att i realtid signalera när en nyhet verkar särskilt viktig och hur det påverkar köp/sälj-läget för en aktie eller annat index.