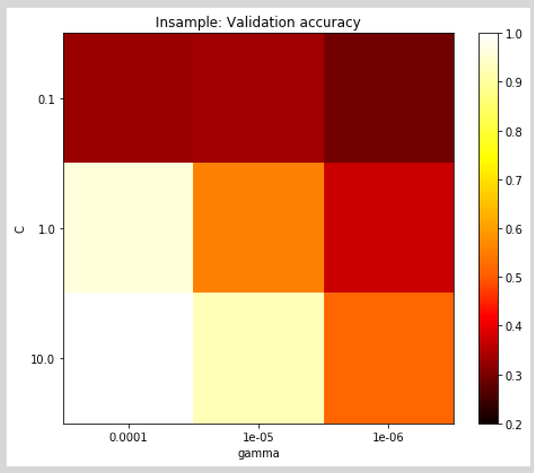
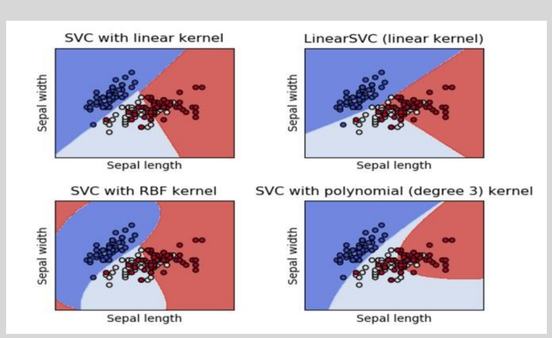
Pioneering the First Course Worldwide on Text Analytics in Finance and Fintech

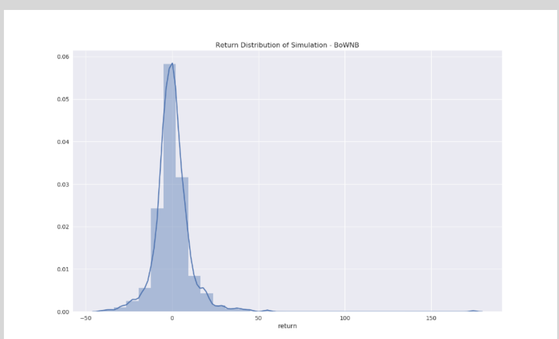
Dr. Matthias Buehlmaier, Faculty of Business and Economics, HKU

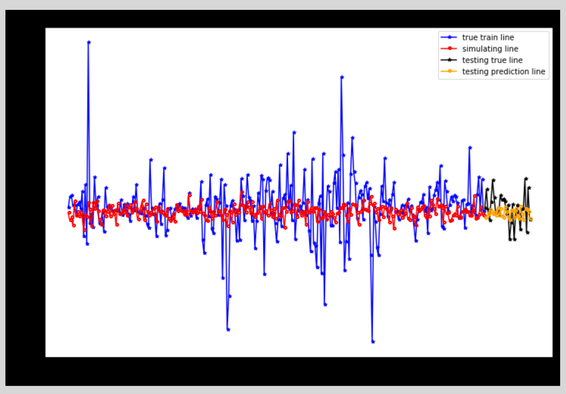
buehl@hku.hk

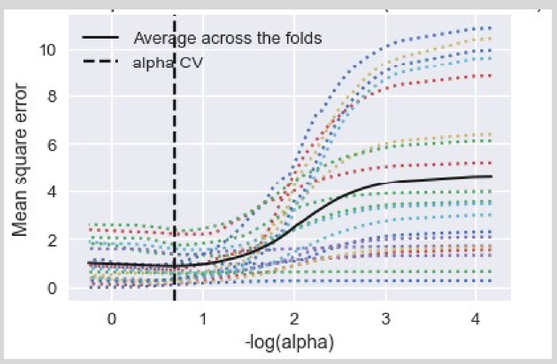
Are Earnings Call Transcripts Informative?









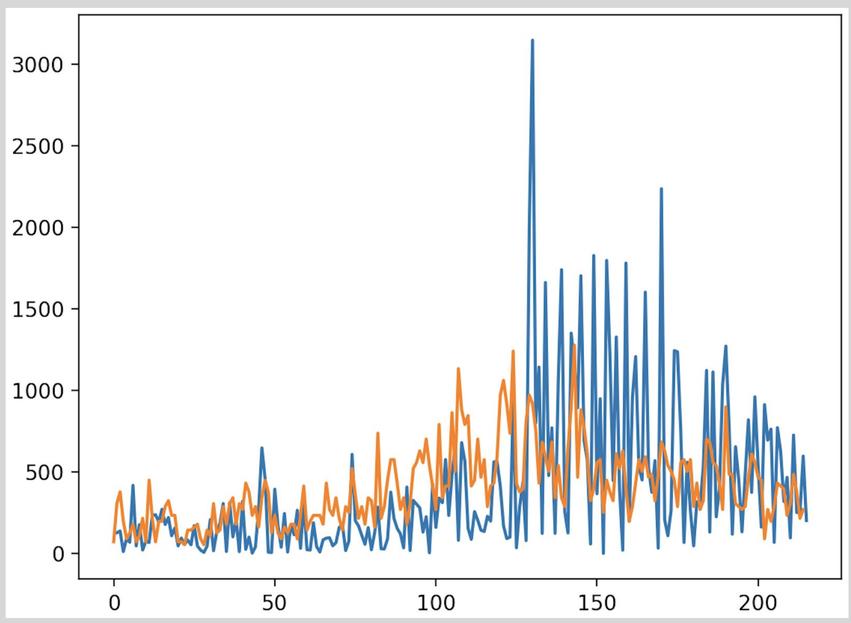


Abstract

This project describes the development and implementation of the first course worldwide on text analytics and natural language processing (NLP) in finance and fintech. This is a cutting-edge topic that is gaining an increasing amount of attention both in academia as well as the industry and with regulators. People often ask me what this is all about. An example I often give is to analyze the words in the written statements by central banks to determine whether they are going to raise interest rates in the future. This is a trillion-dollar question and the techniques learned in this course can provide at least partial answers to this important topic. Another example is to analyze the text in company filings such as annual reports to learn more about the company and to ultimately make more profitable investment decisions. While a lot of work has been done before in fields such as computer science and statistics, there is relatively little prior information available on how this knowledge can be transferred to and applied in fields such as finance and fintech. Based on my own research expertise, this course has been developed to fill this gap and has been successfully taught for the first time worldwide at The University of Hong Kong from January to March 2018 in the Master of Finance Program of the Faculty of Business and Economics. This poster details the development and implementation of this course, which is the first of its kind worldwide.

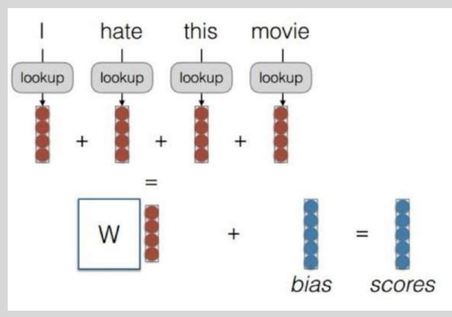
Does Social Media Lead or Lag the Bitcoin Price?

	Dependent Var.	Independent Var.	Coefficient	Adj. R-square	P-level
Matching Dates	% Δ price	sentiment score	0.0005	(0.001)	0.390
	% Δ price	positive ratio	(0.0088)	(0.001)	0.388
	abs(% Δ price)	total tweets	0.0007	0.062	2.13E-08
	abs($ Δ price)	total tweets	15.0644	0.275	4.76E-35
Tweets Lead Bitcoin	% Δ price	score day-1	(0.0003)	(0.002)	0.642
	% Δ price	pos_ratio day-1	(0.0128)	0.002	0.177
	abs(% Δ price)	total tweets day-1	0.0006	0.039	8.62E-06
	abs($ Δ price)	total tweets day-1	14.2196	0.245	8.47E-31
Bitcoin Lead Twitter Discussion	score day+1	% Δ price	7.0522	0.006	0.048
	pos_ratio day+1	% Δ price	0.0247	(0.002)	0.913
	total tweets day+1	abs(% Δ price)	118.7512	0.106	2.16E-13
	total tweets day+1	abs($ Δ price)	0.0205	0.344	2.02E-45

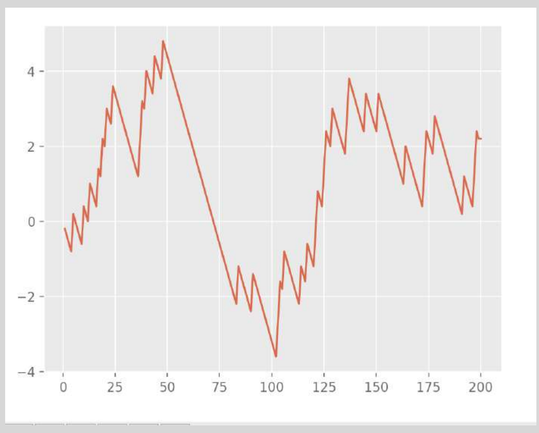


Sentiment Analysis of the U.S. Stock Market


```
from nltk.stem import PorterStemmer
from nltk.tokenize import sent_tokenize, word_tokenize

ps = PorterStemmer()

example_words = ["happiest","happier","happyness","happily"]

for w in example_words:
    print(ps.stem(w))
```




```
def start_stream():
    while True:
        try:
            twitterStream = Stream(auth, listener())
            twitterStream.filter(track=["happy"])
        except:
            continue

start_stream()
```



Predicting Interest Rate Hikes by the Fed Using NLP