

Data Science at Bloomberg

Who are we? Bloomberg's core product, the Terminal, is a must-have for the most influential people in finance. Key participants in the financial markets depend on the Terminal to trade, manage risk, consume information, and navigate time-critical decisions. Our scientists, researchers, and software engineers work on some of the hardest data science problems, especially those related to news analysis and information retrieval.

We develop methods such as semantic and syntactic parsers, NER & NED systems, sentiment analysis, market impact indicators, and social media analysis. Our customers rely on this information to make swift financial decisions; therefore, we need to guarantee precision, accuracy and latency beyond what academia and many other industries would demand.

Our teams: Our core data science team is a close-knit group of 60 engineers focused on NLP, ML, and search.

Natural Language Processing

We are building novel NLP models for name entity recognition, disambiguation and syntactic parsing of social media and other difficult text sources. The infrastructure and algorithms we develop are used widely by data scientists and engineers to build applications such as sentiment analysis of social media (<http://bloom.bg/2litNa0>). We are also building LibNLP: a platform for sophisticated NLP tasks. We leverage open source tools whenever we can and always contribute when we find improvements.

- Work with NLP researchers and other Engineering teams to apply our models to live systems
- Write, test and maintain production-quality C++ code
- Design and evaluate models for use in the fields of data science, AI, machine learning, NLP and information retrieval

Machine Learning (Pattern Recognition)

We are building machine learning models for predicting the impact of news stories on company prices, recommendation systems, semantic parsing for question answering, topic classification, anomaly detection in time series and a variety of other problems at the intersection of structured and unstructured data.

- Design and build systems that solve difficult problems involving text, time series and other complex data sources
- Analyze Bloomberg's unique data to build novel prediction models
- Write, test and maintain production-quality C++ and Python code

Information Retrieval (Relevance and Discovery)

We are building machine learning models to improve user experience via ranking in search and to analyze the connectedness between entities within the financial ecosystem.

- Design and build machine learning systems that solve difficult problems involving knowledge graphs and big data
- Write, test and maintain production-quality Java, C++, and/or Python code
- Work on the main search interface to the Bloomberg Terminal

techatbloomberg.com

 github.com/bloomberg

Bloomberg

Data Science at Bloomberg

How our teams work. Engineers on our Data Science teams have a unique set of skills. They are world-class researchers in NLP, ML, and Search who also know their way around any programming challenge that comes their way. Our teams see their research from conception to production on the Terminal, and learn at each step along the way how to improve their algorithms and systems.

To learn more about the research we're doing, check out some of our recent papers and projects below.

Yi Yang (BBG) and Jacob Eisenstein - [Overcoming Language Variation in Sentiment Analysis with Social Attention](#)

Variation in language is ubiquitous, particularly in newer forms of writing such as social media. Fortunately, variation is not random; it is often linked to social properties of the author. Yi will discuss how scientists can leverage social networks to make sentiment analysis more robust to social language variation. (Presented at ACL, August 2017)

Daniel Preotiu-Pietro (BBG), Jordan Carpenter and Lyle Ungar - [Personality Driven Differences in Paraphrase Preference](#)

Personality plays a decisive role in how people behave in different scenarios, including online social media. Researchers have used such data to study how personality can be predicted from language use. Daniel will discuss how to quantify the predictive power of phrase choice in user profiling and use phrase choice to study psycholinguistic hypotheses. (Presented at ACL, August 2017)

Diego Ceccarelli, Francesco Nidito, Miles Osborne – [Ranking Financial Tweets](#)

This paper, written together by three Bloomberg engineers, examines the many factors that help determine which tweets are relevant for investment decisions. In the future, our engineers will be able to incorporate such innovations into the design of the next generation of machine-learning powered [news rankers](#), ensuring our algorithms will be able to better predict which tweets customers will need, and want, to read.

More:

// Email our recruiter Lauren Tucker: ltucker14@bloomberg.net

// Read about the unique problems we're solving: techatbloomberg.com/post-topic/data-science

// Explore our repo: github.com/Bloomberg

// Follow us on twitter:  @TechAtBloomberg

techatbloomberg.com

 github.com/bloomberg

Bloomberg