In pursuing his degree, Mr. Zou has taken the following relevant coursework: **Statistical Analysis, Mathematical Statistics, Applied Categorical Data Analysis, Modern Applied Statistics, Bayesian Applied Decision Theory, Recent Development in Bayesian Analysis, Statistical and Econometric Methods, and Microeconomics.**

This coursework has prepared Mr. Zou for the Data Scientist position in the following ways:

* **Statistical Analysis:**
  + Hypothesis testing
  + Explorative analysis
* **Mathematical Statistics:**
  + In-depth algorithm
  + Algorithem optimization
  + Mathematics
* **Applied Categorical Data Analysis:**
  + **Frequency**
  + **Customer profile groups**
  + **Risks?**
* **Modern Applied Statistics:**
  + **Python and R**
* **Bayesian Applied Decision Theory:**
  + Expertise and modeling
* **Recent Development in Bayesian Analysis:**
  + Expertise and modeling
  + Recent topics
  + Competitive edge
* **Statistical and Econometric Methods:**
  + **Discrete choice**
  + **Heterogeneity**
* **Microeconomics:**
  + **Customer purchase & behavior & profile**

Credit, PayPal credit, fraud detection, marketing, collections, sales, product, customer lifetime management, customer service

Business intelligence, complex quantitative data mining solutions,

Complex, high-volume, high dimensionality data, including online fraud relational data

risk management model and application platform

>> Building payment fraud detection model with large amount of data  
>> Participating in initial prototyping and implementing of payment network for high capacity and low latency query  
>> Data analysis to understand fraud pattern and model efficiency   
>> Performance analysis of large scale real time data platform and decision engine

Modeling and data analysis: developed end to end (classification and regression) models to rank order and segment customers into different risk groups

Data pipeline: Clean up and prepare massive data for modeling;  
Feature engineering and selection.

Transaction fraud/stolen bank/stolen credit card detection, risk modeling

Transactional data & anomaly detection

User behavior

Account profile

The account engine determines target accounts from profile groups based on behavioral data.

algorithms mine data from the customer’s purchasing history—in addition to reviewing patterns of likely fraud stored in its databases—and can tell whether, for example, the suspect transactions were innocent actions of a globe-hopping pilot.

<https://blogs.wsj.com/riskandcompliance/2015/06/18/how-paypal-manages-fraud-risk/>

The digital payment processor, which split from eBay Inc. in July, uses a combination of technology and human expertise to identify and stop potential fraud in real-time, during online transactions.

The system tags the account for review by human experts, she said. “They might discover that the IP addresses are at airports and this guy is a pilot,” she said. Once verified, that intelligence is fed back into PayPal’s systems. Humans don’t make the system faster, but they make real-time decisions as a check against, and supplement to, the algorithms, she said.

PayPal makes shopping a comfortable task by processing payments of auction websites and vendors through cloud computing, in a safe and secure manner. For PayPal the real consumers are the merchants and every customer of the merchant is indirectly a consumer of PayPal. PayPal provides advanced predictive capabilities to help its merchants improve their customer experience.

Hadoop coexists with traditional data platforms at PayPal to meet various business requirements like customer sentiment analysis, fraud detection and market segmentation.

It collects more than 20 terabytes of log data every day for sentiment analysis, event analytics, customer segmentation, recommendation engine and sending out real-time location based offers.

The data science team at PayPal analyses historical payment data to find out features that indicate an attempted scam. Different types of Machine learning algorithms analyse 1000’s of data points in real- time like - the buying history, recent activity on the merchant’s website or the PayPal site, data stored in cookies, buying history, etc. 300 variables are calculated per event for some of the machine learning models to find a potential fraudulent transaction. The results of analysis are compared with external data provided by authentication providers. For example, if the analysis shows multiple IP addresses from different locations across the globe, for a single account – then it is probably an indication that the account is hacked and it is flagged for review by human experts.

PayPal addresses multi-channel communication happening on tablets, smartphones, in-store and on websites by enticing customers with location based advertisements and offers. As customers today shop in multiple ways – mobile, website and in-store, it is difficult for marketers and advertisers to decide - which is the best bet for placing personalized ads and relevant offers. PayPal is leveraging big data to send relevant customized offers and discounts from merchants to customers. The analytic algorithms use past-purchase history based on the medium of shopping - online or in-app to recommend offers that help customers save money and drive higher transaction volumes for merchants. PayPal incorporates big data analytics to tie customer preferences and tastes, location, purchase history and user activity across various sites, to send relevant offers and discounts along with personalized ads.

PayPal uses data from similar customers to predict the buying behaviour of its customers. The data models look for similar places customers visit. PayPal knows that customers who shop at Home Depot are likely to eat at Subway. Using this analytic insight, PayPal offers discounts on sandwiches at Subway outlets near Home Depot locations.

Transactional data is used to create customer genome sequences which create look-alikes and segment customers into different groups that help create strong signals for personalization, targeted advertising and recommendations. Predictive data models at PayPal make predictions with 69% accuracy on where their customers are likely to spend money.

PayPal’s Hadoop-based text mining system is a critical component for a variety of data science activities at the company, including predictive modeling, sentiment analysis, influence scoring, profile ranking, and topic modeling and clustering.