A purposeful walk down Wall Street
Using advanced data analytics in global financial markets

Presented by Joshua Feinstein
GM IT, WSU College Recruiting Lead
Global Data Strategy, AI and Analytics Services

This book popularized the ‘Random Walk Hypothesis.’ To quote wiki:

“A financial theory stating that stock market prices evolve according to a random walk (so price changes are random) and thus cannot be predicted.”

John Cassidy, from the New Yorker published an article in 2003, citing how Malkiel explained that ‘predicting how the market will behave tomorrow, or next month, isn’t just difficult; it’s impossible.’
Mr. Cassidy’s direct interpretation:

“Financial markets are ‘like a drunk staggering home from the bar, it will lurch one way or another, depending on the laws of chance. To use the language of statistics, the market will follow a ‘random walk.’”
Fast forward a decade:

In the autumn of 2014, GM’s enterprise data professionals teamed with Wayne State University professors to embark on a Big Data and Analytics Capstone Program.

Our ultimate goal was to create an advanced data analytics platform that could predict pricing sentiment for global financial instruments – i.e.

i.e., ‘A purposeful Walk down Wall Street!’
“Prediction is very difficult, especially about the future.” - Neils Bohr, Danish physicist.

The breadth and depth of our undertaking is ever-increasing in scale and has spanned several years.

Our journey has seen us leverage cutting edge data analytics, big data technologies, artificial intelligence and machine learning, as well as mature and traditional programming approaches along the way.
Over the next hour, we’ll share our successes, our challenges, and overall experiences.

We'll look back at our initial concepts and designs. We’ll see how our vision has translated into implementation and evolved over the years.

We will also share what interesting new scenarios have emerged and what is currently in development.
### AGENDA:
Let’s take a purposeful walk through time

<table>
<thead>
<tr>
<th>SEMESTER</th>
<th>FOCUS &amp; DIRECTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014 Fall</td>
<td>An Idea, Geo-Politics, Social Media and Global Market Currencies</td>
</tr>
<tr>
<td>2015 Spring &amp; Fall</td>
<td>Social Media, Twitter &amp; Flume. Taxonomies, Core NLP &amp; Oil.</td>
</tr>
<tr>
<td>2016 Spring &amp; Fall</td>
<td>Neural Networks. ARIMA. Quandl Shift to Quants.</td>
</tr>
<tr>
<td>2018 Spring</td>
<td>Current State of the Platform. Partnership Expansion WSU Grad School Challenge. Neural Networks &amp; Backpropagation</td>
</tr>
</tbody>
</table>
AGENDA: Starting our walk through time

<table>
<thead>
<tr>
<th>SEMESTER</th>
<th>FOCUS &amp; DIRECTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014 Fall</td>
<td>An Idea, Geo-Politics, Social Media and Global Market Currencies</td>
</tr>
<tr>
<td>2015 Spring &amp; Fall</td>
<td>Social Media, Twitter &amp; Flume. Taxonomies, Core NLP &amp; Oil</td>
</tr>
<tr>
<td>2016 Spring &amp; Fall</td>
<td>Neural Networks. ARIMA. Quandl. Shift to Quants</td>
</tr>
<tr>
<td>2018 Spring</td>
<td>Current State of the Platform. Partnership Expansion WSU Grad School Challenge. Neural Networks &amp; Backpropagation</td>
</tr>
</tbody>
</table>
A partnership proposed between GM and WSU to create a real-world project enabling senior WSU students to learn and apply data science, analytics and Big Data technologies. The Use-Case we chose for this project centered around Geo-Politics and Currencies was as follows:

With the modern Geo-Political environment being highly volatile, global market currencies are impacted week-by-week, day-by-day and even hour-by-hour.

We initially proposed to leverage Big Data technologies in an attempt to track and predict the sentiment* of Sovereign Currencies as result of real world events.

* Sentiment is the predictive nature of assessing the future direction of a given currency.
Initial project deliverables

- Identifying of viable data sources.
- Creation of a taxonomy that would capture key words and ideas that could lead to predicting currency sentiment
- Setting up a Hadoop environment
- Applying the taxonomy to create a map-reduce structure in Hadoop
- Creation of load-routines to bring source data into the Hadoop cluster
- Creation of sentiment output for a given currency over a given time period (to be determined during the project life-cycle)

Proposed technologies:

- Hadoop
- Open source ETL tool(s) or other extraction and loading software
- Linux environment preferred. Windows is also acceptable
- Programming and/or scripting language of team’s choice
- AI, Analytics tools to be determined based on team’s choice
Our initial proposed GM-WSU senior project timeline for 2015
### SEMESTER

<table>
<thead>
<tr>
<th>Year</th>
<th>Focus &amp; Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014 Fall</td>
<td>An Idea, Geo-Politics, Social Media and Global Market Currencies</td>
</tr>
<tr>
<td>2015 Spring &amp; Fall</td>
<td>Social Media, Twitter &amp; Flume. Taxonomies, Core NLP &amp; Oil</td>
</tr>
<tr>
<td>2016 Spring &amp; Fall</td>
<td>Neural Networks. ARIMA. Quandl Shift to Quants</td>
</tr>
<tr>
<td>2018 Spring</td>
<td>Current State of the Platform. Partnership Expansion WSU Grad School Challenge. Neural Networks &amp; Backpropagation</td>
</tr>
</tbody>
</table>
2015: Social Media, Twitter, Flume, Core NLP, Taxonomies, & Oil

GM Big Data Currency Sentiment Analysis
Technical Diagram

- Workflow coordination
- Automate processes

- Collects Data using specific keywords
- Stores into HDFS

- Scalable/reliable data storage
- Easy to integrate

- Add Partitions Hourly

- MapReduce
  - Construct complex business logic
  - Spread analysis over multiple clusters
The first semester’s team recommendation:
Tracking the sentiment of Oil, as opposed to Currencies, would be a more accessible initial use-case.
Geo-political conflict and the price of crude oil

The illustration above displays that war in the Middle East causes an increase in the cost of crude oil which reflects in increase of gas prices as well. This upward trend influences the value of currency to rise.
<table>
<thead>
<tr>
<th>SEMESTER</th>
<th>FOCUS &amp; DIRECTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014 Fall</td>
<td>An Idea. Geo-Politics. Social Media and Global Market Currencies</td>
</tr>
<tr>
<td>2015 Spring &amp; Fall</td>
<td>Social Media, Twitter &amp; Flume. Taxonomies, Core NLP &amp; Oil.</td>
</tr>
<tr>
<td>2016 Spring &amp; Fall</td>
<td>Neural Networks. ARIMA. Quandl. Shift to Quants.</td>
</tr>
<tr>
<td>2018 Spring</td>
<td>Current State of the Platform, Backpropagation Neural Networks, Partnership Expansion WSU Grad School.</td>
</tr>
</tbody>
</table>
Neural Networks – What’s the Big Idea?

- Came from neurons in the brain
- Send a signal(s) from one neuron(s) to others
- Neurons receive the signal and decide to fire or not
  - When a signal reaches a certain threshold, they will fire
Autoregressive Integrated Moving Average: ARIMA is a forecasting time series technique that projects the future values of a series based entirely on its own inertia.

https://en.wikipedia.org/wiki/Autoregressive_integrated_moving_average ‘...In time series analysis, an autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model. Both of these models are fitted to time series data either to better understand the data or to predict future points in the series (forecasting).

Moving Average Formula:
Moving Average models account for the possibility between a variable and the residuals from previous periods

\[ y_t = \mu + \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} \]

Autoregressive Formula:
Autoregressive models are models in which the value of a variable in one period is related to its value in the previous period

\[ y_t = \mu + \sum_{i=1}^{p} Y_i y_{t-i} + \epsilon_t \]
Quandl: https://en.wikipedia.org/wiki/Quandl

Quandl (ˈkwændəl/) is a platform for financial, economic, and alternative data that serves investment professionals. Quandl sources data from over 500 publishers.\[1\] All Quandl's data are accessible via an API.\[2\] API access is possible through packages for multiple programming languages including R, Python, Matlab, Maple and Stata.
2016: Shifting to Quants
2016: Neural Networks. ARIMA. Quandl. Quants.
<table>
<thead>
<tr>
<th>SEMESTER</th>
<th>FOCUS &amp; DIRECTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014 Fall</td>
<td>An Idea, Geo-Politics, Social Media and Global Market Currencies</td>
</tr>
<tr>
<td>2015 Spring &amp; Fall</td>
<td>Social Media, Twitter &amp; Flume. Taxonomies, Core NLP &amp; Oil.</td>
</tr>
<tr>
<td>2016 Spring &amp; Fall</td>
<td>Neural Networks. ARIMA. Quandl Shift to Quants.</td>
</tr>
<tr>
<td>2018 Spring</td>
<td>Current State of the Platform. Partnership Expansion WSU Grad School Challenge. Neural Networks &amp; Backpropagation</td>
</tr>
</tbody>
</table>
2017: A Break Through With Neural Networks

Previous Work and State of FASA
Other approaches utilize different levels of sentiment analysis to make predictions as well as many different neural network configurations to make the financial prediction. The previous semesters utilized a single neural network combined with sentiment analysis to make their predictions.

Technical Objectives
Signal processing techniques and multiple neural networks were employed to improve the overall accuracy of the predictions. The classic ARIMA algorithm was used as a baseline metric for the machine learning approach.

Technical Approach and Results
Signal processing algorithms were employed to preprocess the data used as input into our array of neural networks. By employing this preprocessing technique and using multiple neural networks, students were able to increase the prediction accuracy by over 20% for all metrics. Metrics and trend lines were added to a simple dashboard style interface to allow for visualization of the data.

Next Steps for Development and Test
Further development is required to open up the application to new data sources. In particular, neural networks could be used to find trading signals within a financial instruments time series instead of the more classic financial indicators. The weight vectors of this neural network could help shed light on which variables are most important.
## Continuing our walk through time

<table>
<thead>
<tr>
<th>SEMESTER</th>
<th>FOCUS &amp; DI RECTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014 Fall</td>
<td>An Idea, Geo-Politics, Social Media and Global Market Currencies</td>
</tr>
<tr>
<td>2015 Spring &amp; Fall</td>
<td>Social Media, Twitter &amp; Flume. Taxonomies, Core NLP &amp; Oil.</td>
</tr>
<tr>
<td>2016 Spring &amp; Fall</td>
<td>Neural Networks. ARIMA. Quandl Shift to Quants.</td>
</tr>
<tr>
<td>2018 Spring</td>
<td>Current State of the Platform. Partnership Expansion WSU Grad School Challenge. Neural Networks &amp; Backpropagation</td>
</tr>
</tbody>
</table>
2018: Current State of the Platform

NEURAL NETWORKS TRACKING COMMODITIES (OIL)

~$64 per barrel as of 2/26

$56.73 on 6/30/18

$56.73 on 6/30/18
2018: Current State of the Platform

NEURAL NETWORKS & THE CROSS MOVING AVERAGE TRADING STRATEGY TRACKING COMMODITIES (OIL)
2018: Current State of the Platform

NEURAL NETWORKS TRACKING COMMODITIES (GOLD)

Gold down ~$7 on 6/30/18
AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) & THE CROSS MOVING AVERAGE TRADING STRATEGY TRACKING COMMODITIES (GOLD)
NEURAL NETWORKS AND THE CROSS MOVING AVERAGE TRADING STRATEGY TRACKING INDICES (S&P 500)
The FASA (Financial Analytics and Sentiment Analysis) System currently tracks:

- Commodities
- Currencies
- Equities and Indices

At present, all examples reviewed are run separately with separate outputs.

Through partnering with WSU Business Analytics Graduate students, we are now endeavoring to combine both cutting-edge AI Neural Networks with traditional Trading Strategies.

Our hypothesis is that, by combining two or more methodologies together, the results will be far more superior and accurate.

Of course, we are simultaneously working with the graduate and undergraduate students to improve the individual AI algorithms.
2018: Current Application Flow

- Prediction Server: Prediction Source, Horizon, Date Range
- Quandl API: Quandl Code for Financial Instrument, Horizon, Date Range, API Authentication String
- MySQL Database: Insert Predicted Values
- Predictions Engine: Horizon, Prediction Type, Financial Instrument, Database Credentials, Column Names, Date Range
- PredictorN: Historical Dataset
- Neural Network: Model File Path, Number of Nodes, Number of Networks, Lockback, Horizon
- NNensemble
- Save to Database: Table Name, Horizon, Prediction Dataset
- Prediction Outframe: Quandl Historical Dataset, Financial Instrument, Horizon, Column Names, Date Range
A word about General Motors

- **Top global automaker**
  - Fortune 8 company
  - 2017 sales of 9 million vehicles and revenue >USD$145.6B
  - ~181,000 employees globally
  - Our goal: to earn customers for life

- **GM by the numbers**
  - Produce vehicles in 30 countries; do business in 125 countries
  - 12,450 dealers
  - 7 million OnStar (4G LTE) subscribers
  - >170 vehicle assembly, metal stamping, tool and die, propulsion, components and battery assembly manufacturing facilities and 20 customer care & aftersales (CCA) facilities
  - 65,000 tons of materials and 29,000 outbound vehicles shipped daily

- **GM's average vehicle has 100 million lines of code**
This is GM: Chevrolet, Cadillac, GMC, Buick

2018 Chevrolet Camaro ZL1

2018 Cadillac CT6

2018 Buick Enclave

2018 GMC Acadia Denali
A global company

Holden Captiva 7

Baojun 730

Wuling Hongguang S3

Jefang F330
Roaring into No. 1 for her third straight year, Barra remains in the driver’s seat

- Most powerful woman (Fortune magazine), three years running
- First woman to top Motor Trend’s 2018 Power List, named Person of the Year
- CEO and Chairman of GM since 2014, leading a strong and competitive senior leadership team
  - Corporate officers bring more than 250 years of GM experience to the table, with 70 percent seasoned GM veterans and 30 percent externally-experienced talent
  - Board of directors is diverse; 45 percent female
We are in the midst of a transportation revolution.

“The American automotive giant announced that it is working toward an all-electric, zero-emissions future. That starts with two new, fully electric models next year—then at least 18 more by 2023.”
GM makes IT outsourcing a thing of the past

Pharmaceutical giant AstraZeneca is reversing a decades-old practice of outsourcing its IT, and joining the growing ranks of big corporations that are developing more technology in-house, a shift that they view as key to staying competitive.

The Morning Download: Target CIO, in Shift from IT Outsourcing, Hires Software Developers

GM Builds Upon IT Transformation With New Investments

AstraZeneca Latest Big Company to 'Insourse' IT

The Wall Street Journal

AstraZeneca is reversing a outsourcing its IT, and joining corporations that are technology in-house, a shift that they competitive.

MIT Technology Review

Business Impact

With Computerized Cars Ahead, GM Puts IT Outsourcing in the Rearview Mirror

The Detroit automaker is bringing thousands of IT jobs back in-house as it seeks faster software innovation.

by Jessica Leber  November 5, 2012

CIO Randy Mott has moved the car maker in a similar direction. "To transform the company, you really need IT, which touches all parts of the business."
Innovation through learning and competition

Hackathons are held annually in each Innovation Center, as well as other fun events: Last Idea Standing, Gears for Geeks, Innovation Challenge, Top Coder, Geek Bowl.
GM’s Maxis: Maximizing insights

- Advanced search engine of GM and third-party data, launched in 2017
- Focus on vehicle, plant, customer, dealer, part, supplier data analysis
- Used to project warranty or recall costs, break-even points, or market and vehicle segment trends
- Contains a unique social component that allows users to see who's looking at similar data for advice or discussion
A very special thanks to Professors Sam Bryfczynski and Khayyam Hashmi

Suggested reading, articles and presentations on financial data systems:

2) Database in Depth: Relational Theory for Practitioners by C. J. Date. Published, May, 2005
3) Date on Database: Writings 2000-2006 by C. J. Date. Published Nov, 2006
5) The Global Data Initiative: Enterprise Data World at the DAMA / Wilshire annual conference, Tampa, Florida; April 11, 2009 [link]
6) All Roads Lead to Rome -The Hybrid RDMBS / XML Database at the Enterprise Data World Annual Conference (DAMA / Wilshire), San Francisco, California; March 14, 2010 [link]
8) Practical Applications of Oracle XML DB: Standard and Poor's Use Case at the Oracle Open-World Annual Conference, San Francisco, California; September 22, 2010 [link]
BACK-UP SLIDES
Pure ARIMA

\[ Y_t = c + \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \theta_1 e_{t-1} + \ldots + \theta_q e_{t-q} + e_t \]

Where:

- \( Y_t \) = the variable that will be explained in time \( t \);
- \( c \) = constant or intercept;
- \( \phi \) = coefficient of each parameter \( p \);
- \( \theta \) = coefficient of each parameter \( q \);
- \( e_t \) = Residuals or errors in time \( t \).