

# Jupyter Notebooks: Interactive Visualization Approaches

**Bloomberg**

**QCon 2018  
San Francisco  
November 07, 2018**

**Chakri Cherukuri  
Senior Researcher  
Quantitative Financial Research Group**

**TechAtBloomberg.com**

# Outline

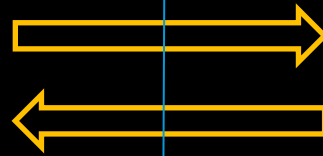
- Introduction
- Overview of interactive widgets
- Example with code walk-through
  - Normal Distribution
- Case studies
  - Server log dashboard
  - Twitter sentiment analysis
  - Tools for deep learning
- Q & A

# Interactive widgets

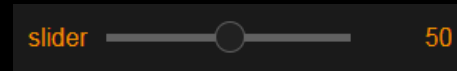
## Python Object

```
IntSlider(description='slider',  
value=50)
```

- Interface the user sees
- Its attributes (*traits*) can send events when updated



## Visual Representation



- Graphics, Interaction
- JavaScript implementation hidden from the user

# Interactive widgets

- **ipywidgets**: core UI controls (text boxes, sliders, button etc.)
- **bqplot**: 2D plotting widgets (built on top of the ipywidgets framework)
- **Key takeaways:**
  - **ipywidgets** and **bqplot** can be combined to build interactive visualizations
  - Attributes of widgets (traits) can be linked using callbacks





# Server Logs Dashboard

**TechAtBloomberg.com**

**Bloomberg**

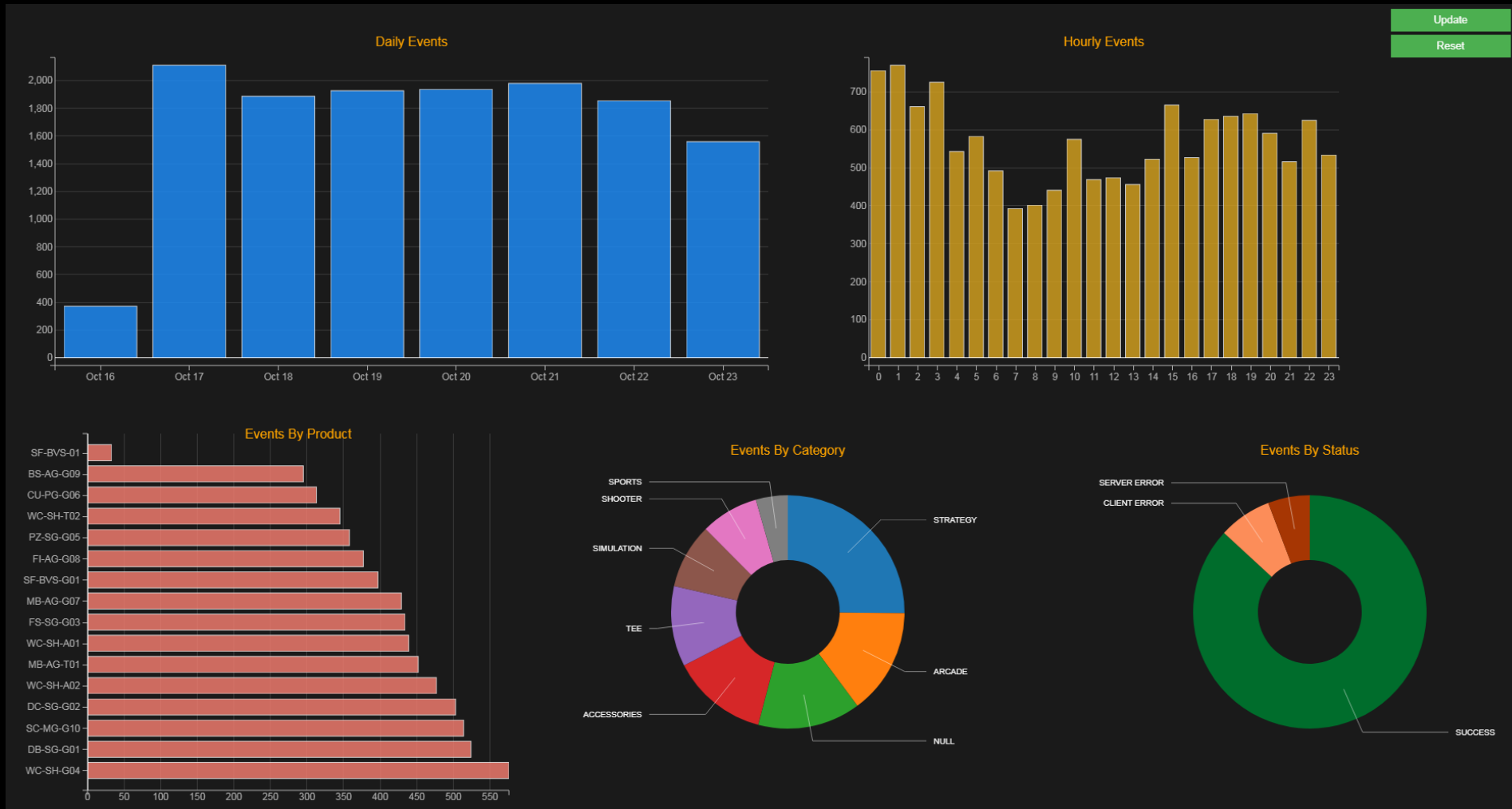
# Server Logs

- Information we can glean:
  - Timestamp of requests/events
  - IP address
  - HTTP status codes
  - Agent
  - URL
- URLs can be parsed to obtain search parameters:
  - Product IDs
  - Category IDS

# Dashboard Components

- Time series plots of events
  - Trends
  - Seasonality
  - Outliers
- Daily/hourly aggregation of events
- Plots of events broken down by
  - Products
  - Categories
  - Status codes

# Server log dashboard



TechAtBloomberg.com

Bloomberg



# Performance/Scaling

- Data ingestion from streaming sources like **Kafka**
- Notebook servers running on spark/hadoop clusters
- Streaming data frames
  - dask dataframes
  - pandas with chunking



# Twitter Sentiment Analysis

**TechAtBloomberg.com**

**Bloomberg**

# News/Twitter Sentiment

- Social sentiment from the raw news story or tweet. Data is:
  - Unstructured
  - Highly time-sensitive
- Story-level sentiment
- Company-level sentiment
- Sentiment score can be used as a trading signal

# Twitter Sentiment Classification

## Problem statement:

Predict the sentiment (negative, neutral, positive) of a tweet for a company

Ex: “\$CTIC Rated strong buy by three WS analysts. Increased target from \$5 to \$8.” = Positive

Three way classification problem

- Input: raw tweets
- Output: sentiment label  $\sum$  {negative, neutral, positive}

# Methodology

- We are given **labeled** training and test data sets
- Train classifier on training data set
- Predict labels on test data and evaluate performance



# One vs. rest Logistic Regression

- Train three binary classifiers for each label
  - Model 1: Negative vs. Not Negative
  - Model 2: Positive vs. Not Positive
  - Model 3: Neutral vs. Not Neutral
- Get probabilities (measures of confidence) for each label
- Output the label associated with the highest probability

# Classifier performance analysis

- Look at misclassifications
  - Confusion Matrix
- Understand model predicted probabilities
  - Triangle visualization
- Fix data issues

# Confusion Matrix

K x K matrix where K = number of classes

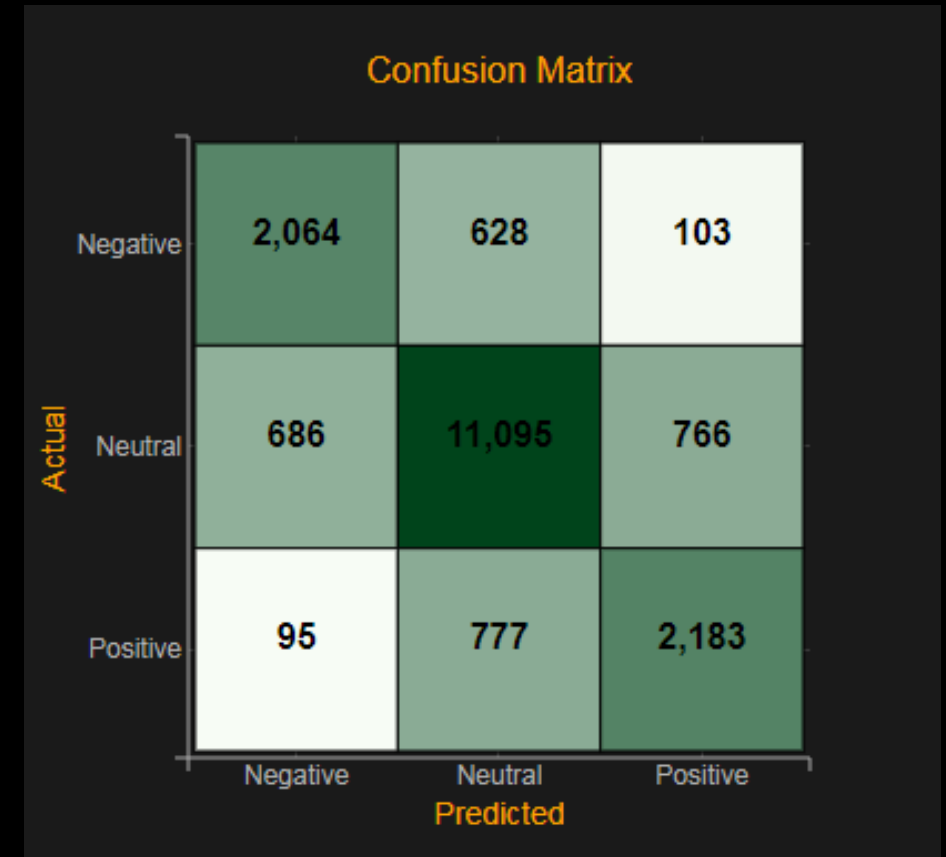
Cell[i, j] = number of samples whose:

Actual label = i

Predicted label = j

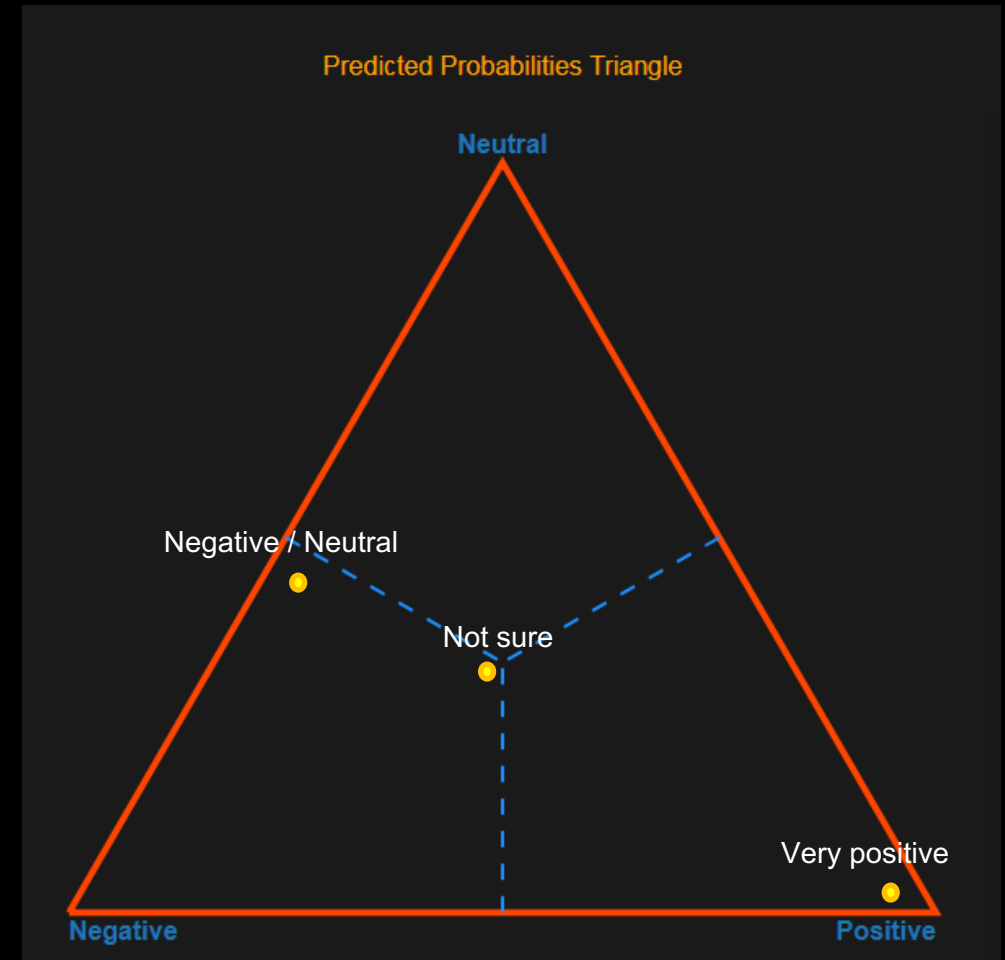
Diagonal entries: **correct predictions**

Off diagonal entries: **misclassifications**



# Triangle Visualization

- Model returns 3 probabilities (which sum to 1)
- How can we visualize these 3 numbers?
  - Points inside an equilateral triangle



# Performance analysis dashboard

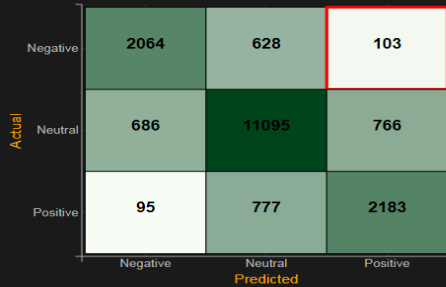
Use the dashboard to:

- Analyze misclassifications (using confusion matrix)
- Improve model by adding more features (by looking at model coefficients)
- Fix data issues (using triangle and lasso)



# Analyze Misclassifications

Confusion Matrix

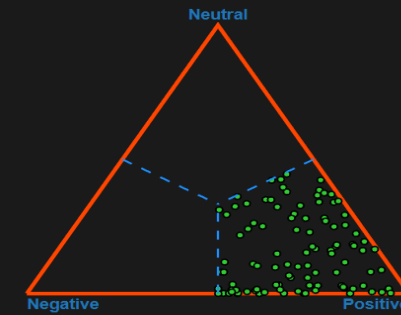


Actual: Negative, Predicted: Positive (103 tweets)

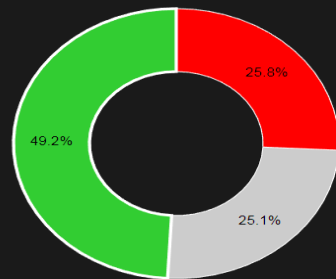
i	Tweet
177	\$PLUG Good day, but still off \$1.20 from the last ER run very sad performance for a company with such potential.
182	\$MAR up 1.0% to 73.50 while \$HOT slips 3.0% to 72.75 in PreM on Marriott to acquire Starwood Hotels in 12.2B Dea
311	\$SMG SMG Stock year to date has changed -.63% percent. +4.37% in the last 30 days. \$SMG Scotts Miracle-Gro C
423	The Wall Street Journal: Starbucks sales growth disappoints: Coffee giant posts 10% revenue gain, divr.it/7Mrkb2 MA.
699	Weight Watchers Sees Profit Slimmed 37% divr.it/7MGwvN
729	\$IOSP IO SP Stock year to date has changed -2.73% percent. +11.37% in the last 30 days. \$IO SP Innospec Inc Full b
757	\$XLE was looking great earlier. Emphasis on "was"...
890	\$GRPN ER results not looking good
1073	JACK DORSEY HAS DESTROYED NEARLY \$5 BILLION IN VALUE IN THE PAST 11 DAYS – Trading with The Fly \$
1141	\$SPY rate hike coming, dollar strong, not looking good for corp earnings going forward and buybacks!
1316	\$SPY A socialist win moves the market down. New job offers for marijuana bike couriers hits new high
1412	business: Siemens announces share buyback and expects no growth in 2016 profit margin bloom.bg/1SkEsqZ pic.twt
1422	\$NURO bear market but with recent highlight this tock is good
1435	Argos Christmas Sales Down Despite Digital Boost: Home Retail Group, the owner of Argos, called the results 'm... n.

Selected Tweet: \$GRPN ER results not looking good

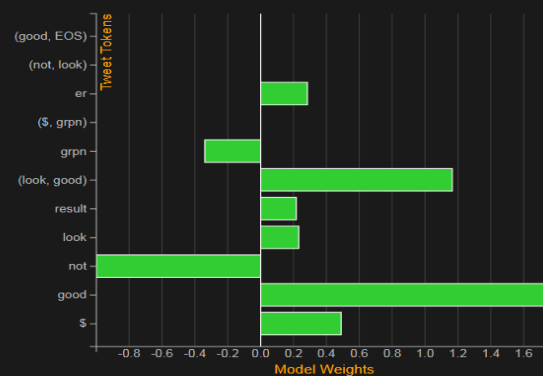
Predicted Probabilities Triangle



Predicted Probabilities

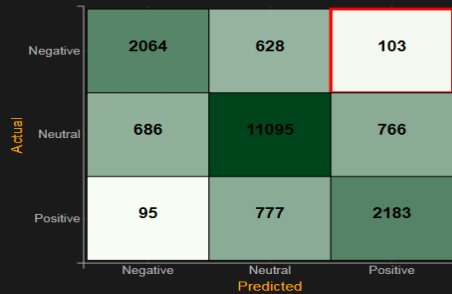


Features & Model Weights



# Analyze Misclassifications

Confusion Matrix

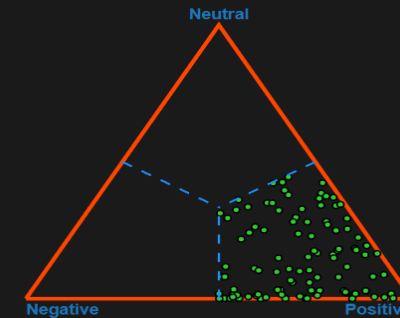


Actual: Negative, Predicted: Positive (103 tweets)

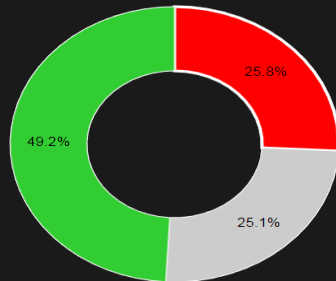
i	Tweet
177	\$PLUG Good day, but still off \$1.20 from the last ER run very sad performance for a company with such potential.
182	\$MAR up 1.0% to 73.50 while \$HOT slips 3.0% to 72.75 in PreM on Marriott to acquire Starwood Hotels in 12.2B Dea
311	\$SMG SMG Stock year to date has changed -0.63% percent. +4.37% in the last 30 days. \$SMG Scotts Miracle-Gro C
423	The Wall Street Journal: Starbucks sales growth disappoints: Coffee giant posts 10% revenue gain, dlvr.it/7Mrkb2 MA.
699	Weight Watchers Sees Profit Slimmed 37% dlvr.it/7MGwvN
729	\$IOSP IO SP Stock year to date has changed -2.73% percent. +11.37% in the last 30 days. \$IO SP Innospec Inc Full b
757	\$XLE was looking great earlier. Emphasis on "was"...
890	\$GRPN ER results not looking good
1073	JACK DORSEY HAS DESTROYED NEARLY \$5 BILLION IN VALUE IN THE PAST 11 DAYS – Trading with The Fly \$
1141	\$SPY rate hike coming, dollar strong, not looking good for corp earnings going forward and buybacks!
1316	\$SPY A socialist win moves the market down. New job offers for marijuana bike couriers hits new high
1412	business: Siemens announces share buyback and expects no growth in 2016 profit margin bloom.bg/1SkEsqZ pic.twit
1422	\$NURO bear market but with recent highlight this tock is good
1435	Argos Christmas Sales Down Despite Digital Boost: Home Retail Group, the owner of Argos, called the results 'm... n.

Selected Tweet: \$GRPN ER results not looking good

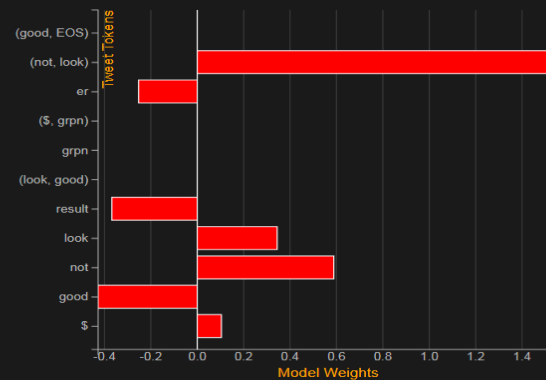
Predicted Probabilities Triangle



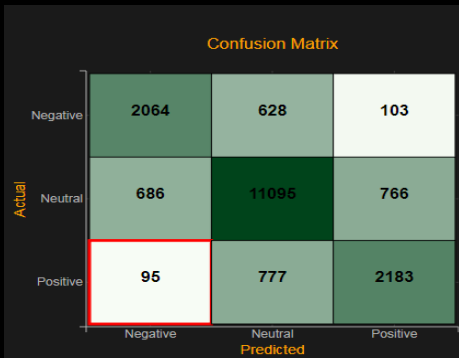
Predicted Probabilities



Features & Model Weights



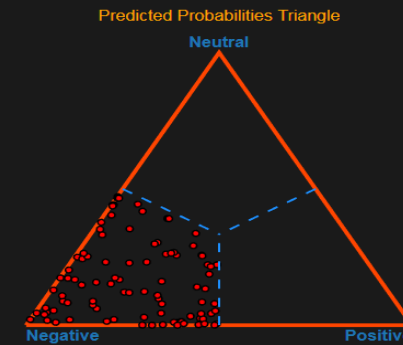
# Analyze Misclassifications



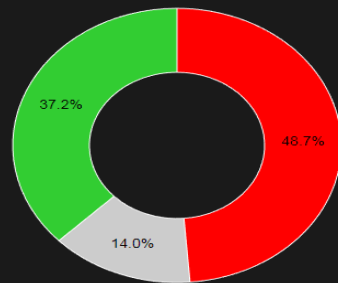
Actual: Positive, Predicted: Negative (95 tweets)

i	Tweet
422	\$SGYP open tells me cash overhang was even greater than I thought. I'm very surprised, very happy at reaction. The
664	\$FB all time highs as expected... from mid Oct lows draw a support line will be nice add spot on any dips imo
883	\$AGN fing " News Algo" picking up onld stories and taking this down, LARGE BUYER
1122	\$SWIR Good company. Bad shortles. Squeeze 'em! http://stks.co/q361z
1426	Anyone else notice the sudden rise in #SolarCity stock off its lows? @elonmusk buy a whole bunch more again? #sct.
1460	Microsoft Tops Exxon as 2nd Biggest Company on Oil Drop - Bloomberg via @BloombergNews bloom.bg/1pW3eow
1733	T-Mobile reaches \$90 million settlement with feds over cramming if.t/16ApAU1
1855	Apple iPhone Sales, Up 35%, Disappoint on.wsj.com/1GAQe95
1971	\$RICK Cash machine. Will bounce back. Surprised it stayed this low for this long.
2017	@SmCapStrategist \$DAR up 2% after weak earnings.
2135	\$PSEC CONFERENCE CALL >FANTASTIC< All must listen This stock will be over \$ 9.00 by years end. Funds will st.
2210	The Guardian: Petrofac Issues profit warning. buff.ly/1xrU6Fx
2424	SPX up 1.7% Nasdaq up 2% DJIA up 1.4% WTI crude up 2.3% (31.15) Comex gold down 0.9% Euro 1.0864 Yen 118.
2454	\$FCX weak rally = great short. been on watch list for a long time.

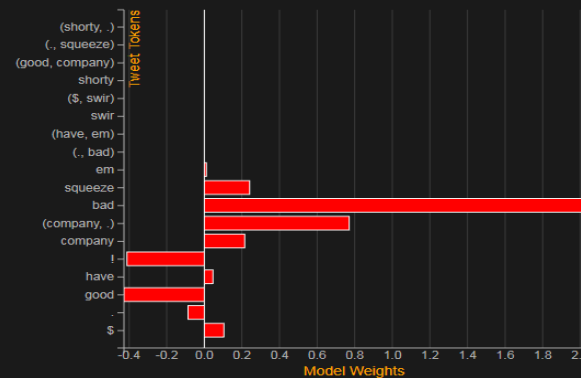
Selected Tweet: \$SWIR Good company. Bad shortles. Squeeze 'em! http://stks.co/q361z



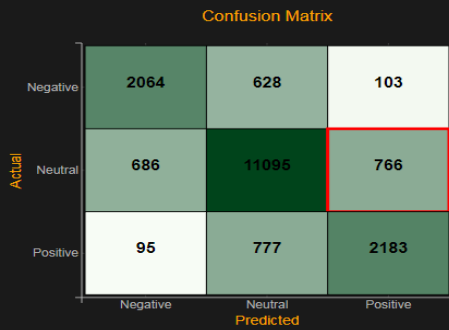
Predicted Probabilities



Features & Model Weights



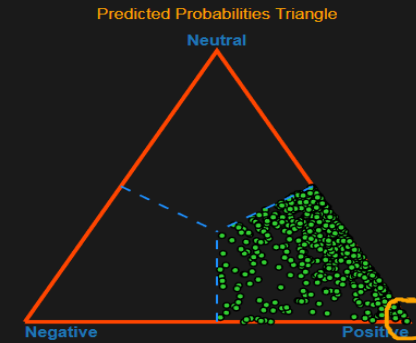
# Use Lasso To Find Data Issues



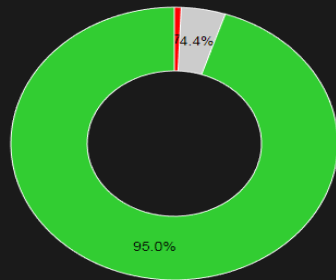
Actual: Neutral, Predicted: Positive (766 tweets) (Lasso Selection: 23 tweets)

I	Tweet
1950	Legg Mason upgraded by Zacks Investment Research to hold. <a href="#">ift.tt/1H75X7J</a> \$LM #LM
2199	UPDATE 2-Microsoft sales beat Street hopes, cloud profits up: SEATTLE, Oct 23 (Reuters) - Microsoft Corp repor... r...
4041	DTN Stock Market: Razen to spend \$1bn on ethanol boost: Royal Dutch Shell-Cozan joint venture to invest in pl... on.
4101	Allstate profit beats estimates as premiums rise: (Reuters) - Allstate Corp , the largest publicly traded home... bit.ly/1D
5945	ALERT: New William %R signal created on \$VOD - Direction: High at 212.55 - Time: 15:05 - Strength: 7.
5964	US Markets at open: DOW (\$DJIA) +0.12%   NASDAQ (\$COMPX) +0.11%   S&P 500 (\$SPX) +0.04%   More: <a href="#">htt</a>
6178	My guess is that Einhorn is wrong on this one and over the next year \$X continues to surprise on upside bloomberg.c.
8684	\$COP COP Gapped up +0.05 \$COP Last Price 73.15 Day High 73.22 Day Low 72.81 Change +0.99 Volume 501,780.
9416	\$XOMA im guessing this is worth min 8\$ ...and whoever is upgrading is buying in...
9435	\$ARMH ARMH Stock Gapped up +0.18 \$ARMH Last Price 44.64 Day High 44.66 Day Low 43.99 Change +1.17 Volu.
10454	RT @jose_garde: Wal-Mart earnings beat expectations; shares up - <a href="#">klou.tt/1tp6cxayso6ug</a> #ecommerce
10779	Mobile Mini PT Raised to \$49.00 \$MINI <a href="#">ift.tt/1CAGKUp</a>
10961	Mirati Therapeutics upgraded by Zacks to buy. \$47.00 PT. <a href="#">ift.tt/1PnlzVB</a> \$MRTX #MRTX
11160	DTN USA: Starbucks: Prices for some drinks to go up by 5 to 20 cents: The Seattle-based company also raised pr... f.

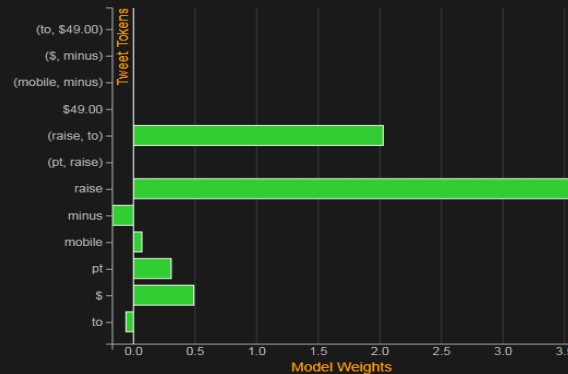
Selected Tweet: Mobile Mini PT Raised to \$49.00 \$MINI [ift.tt/1CAGKUp](#)



**Predicted Probabilities**

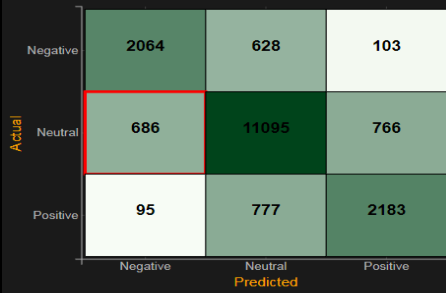


**Features & Model Weights**



# Use Lasso To Find Data Issues

Confusion Matrix

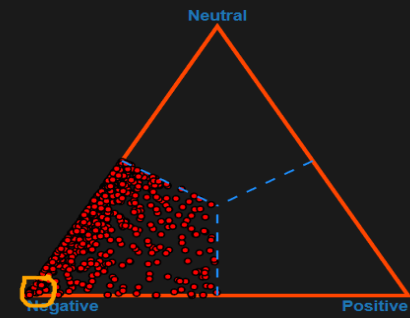


Actual: Neutral, Predicted: Negative (686 tweets) (Lasso Selection: 27 tweets)

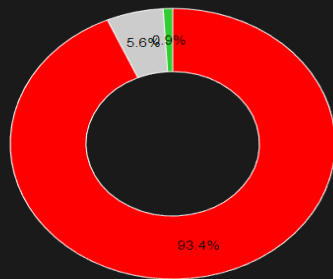
i	Tweet
2366	Amazon Is Twisting The Knife While Tesco Struggles With Accounting Scandal divr.it/7TFFmx via @businessinsider
2943	@biotechpro \$XOMA Great Endocrine Portfolio! http://stks.co/q35ws&quot; POS 1 NX6USJBE07IC Visual Studio 201
4909	Electrolux CEO steps down after failed takeover of GE unit #seattle did/bz/egZyk
4989	H & R Block Price Target Cut to \$37.00 by Analysts at Credit Suisse \$HRB ift.t/1zv3tsl #acn
5510	Apple shares drop off report of weak iPhone demand. Report: Credit Suisse says the company is slashing compon...
5965	DTN USA: BP's \$18 billion spill settlement: BP agreed to pay \$18.7 billion in compensation for the Gulf oil sp... nbcne.
6487	\$UVXY rally tomorrow, or do things really get cooking? If you sell now, you may miss the gap up. OTOH, rally could m
6721	JP Morgan facing currency probe: JP Morgan Chase, the largest bank in the US, has revealed it is under investi... bbc
6773	#News : SHUT IT DOWN Sony reportedly halts production amid hack attack: fxn.ws/1zKYjsg via @TheNewsHype
7759	@twistools_en Pity about emission cheating but nothing wrong with VW engineering. Good cars.
8954	RT @rook429: @MarcACaputo Jesus promoted wine over coffee, Starbucks sent Judas back in time to betray Jesus.
10557	\$RATE faces shareholder lawsuit re accounting errors. #CFO demoted. #GAAP #AUDIT bit.ly/1oDvwfw
10892	6 Reasons Apple Could Have Fallen 6% Yesterday: What makes stocks go up and down? I have no idea -- which ma.
11380	\$HLF-Ackman shall trying to pass off a Herbalife weight loss aide as a caffeine tablet, forgets the herbal blend. -\$HLF

Selected Tweet: Electrolux CEO steps down after failed takeover of GE unit #seattle did/bz/egZyk

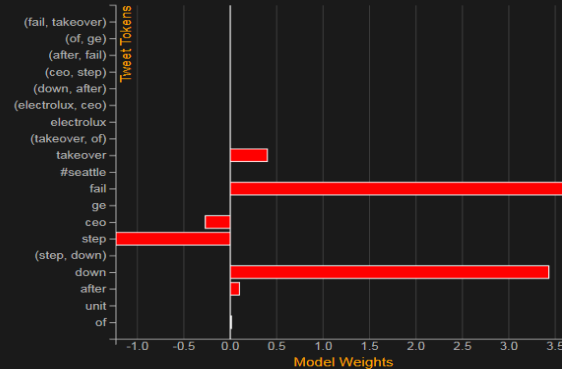
Predicted Probabilities Triangle



Predicted Probabilities



Features & Model Weights







# Tools for Deep Learning

**TechAtBloomberg.com**

**Bloomberg**

# Deep Learning Tools

## Graphical Wizard

- Select network parameters
- Build network architecture
- Training plots
  - Real time loss/accuracy curves
  - Distribution of weights/biases/activations
- Diagnostic plots
  - Residual vs. Predicted Values
  - Confusion Matrix

# Network Parameters

Network Parameters | Architecture | Training | Diagnostics

Epochs	100
Batch Size	64
Loss	mse
Optimizer	adam

optimizer params

lr	0.001
beta_1	0.9
beta_2	0.999
epsilon	1e-8
decay	0

# Network Architecture

Network Parameters | **Architecture** | Training | Diagnostics

Hidden Layers + -

	Inputs	Hidden Layer 1	Hidden Layer 2	Hidden Layer 3	Hidden Layer 4	Outputs
<b>Nodes</b>	4	100	60	30	10	1
<b>Activation</b>		relu ▾	relu ▾	relu ▾	relu ▾	linear ▾
<b>Batch Norm</b>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
<b>Dropout prob</b>		0	0	0	0	

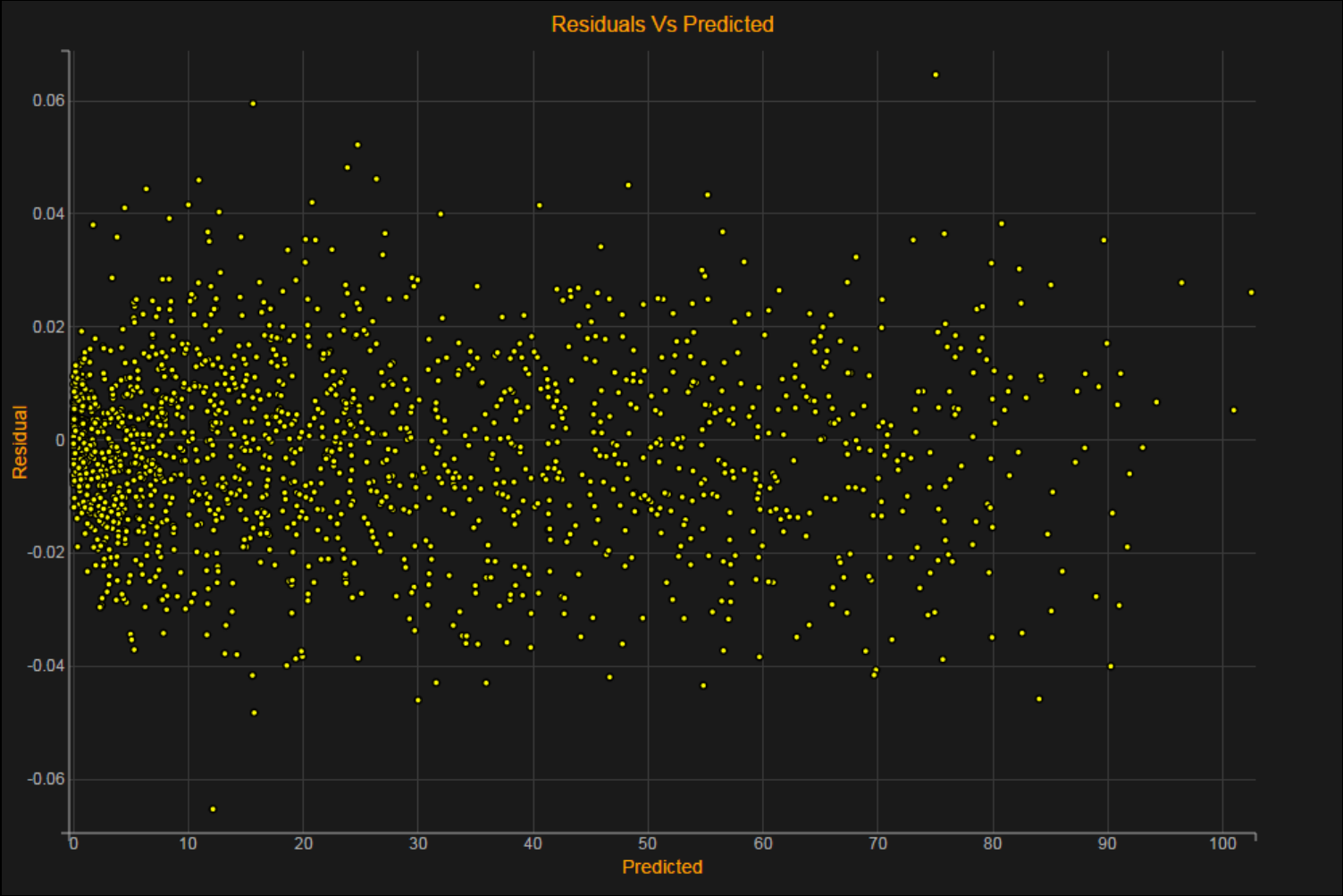
# Loss/Accuracy Curves



# Distributions of Weights/Biases/Activations



# Diagnostic Plots



TechAtBloomberg.com

Bloomberg



# Resources

- Widget libraries used to build the applications:  
**ipywidgets:** <https://github.com/jupyter-widgets/ipywidgets>  
**bqplot:** <https://github.com/bloomberg/bqplot>
- Machine Learning libraries  
**scikit-learn:** <https://http://scikit-learn.org>  
**tensorflow:** <https://www.tensorflow.org>  
**keras:** <https://keras.io>
- Link to the notebooks/code: [https://github.com/ChakriCherukuri/qcon\\_2018](https://github.com/ChakriCherukuri/qcon_2018)
- Tech At Bloomberg blog: [www.TechAtBloomberg.com](http://www.TechAtBloomberg.com)

