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

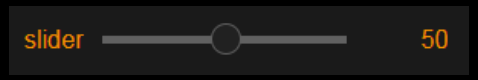
Interactive widgets

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

TechAtBloomberg.com

Bloomberg

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

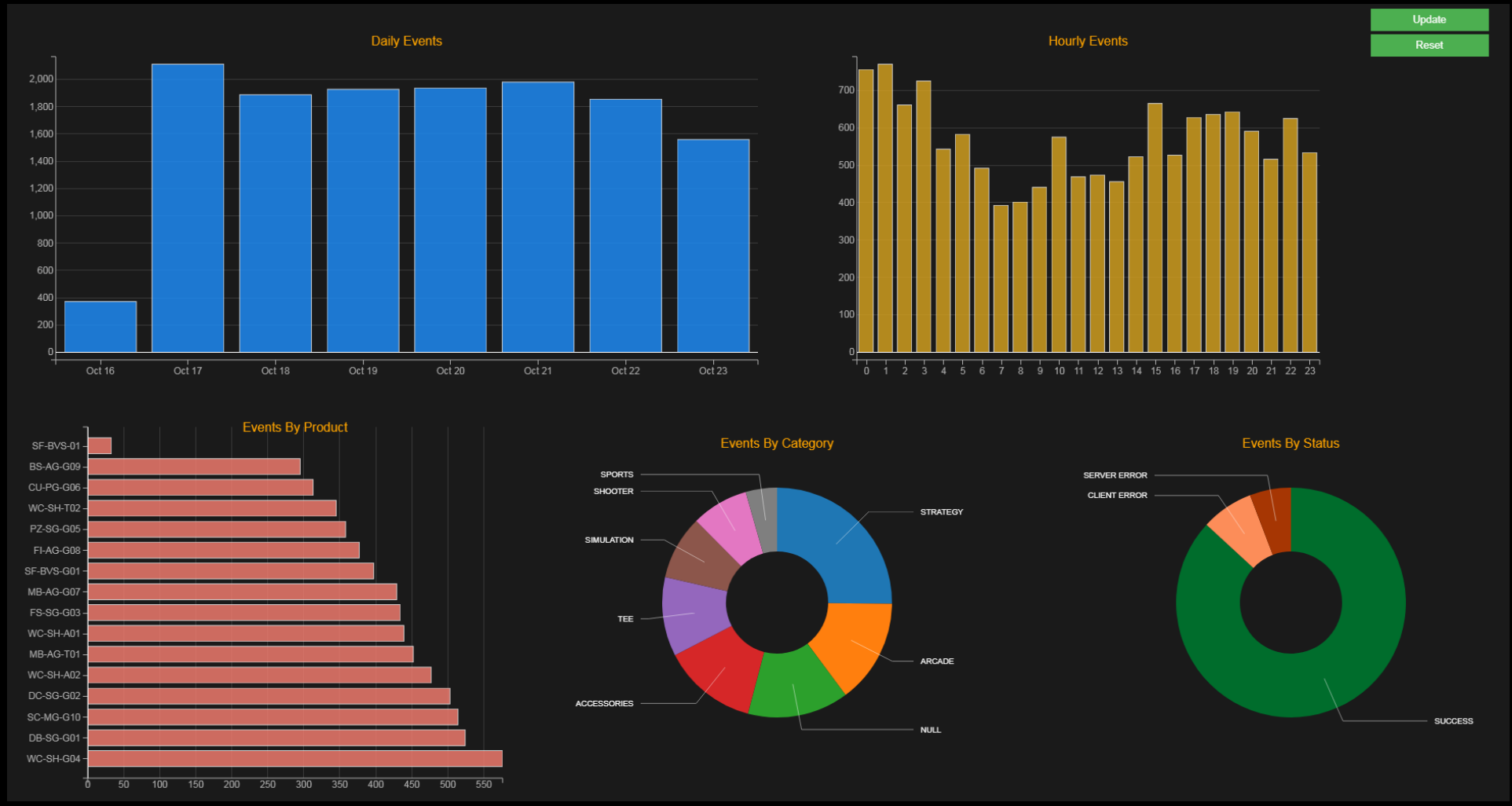
Dashboard Components

- 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 ∑ {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

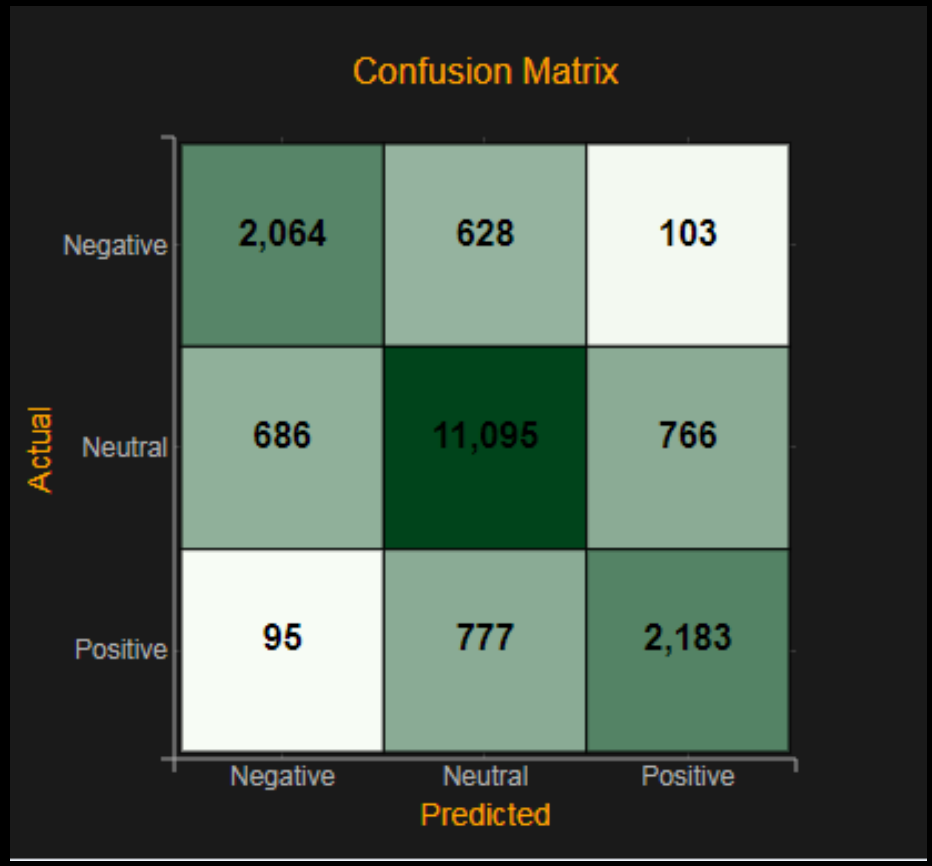
Cell[i, j] = number of samples whose:

Actual label = i

Predicted label = j

Diagonal entries: correct predictions

Off diagonal entries: misclassifications



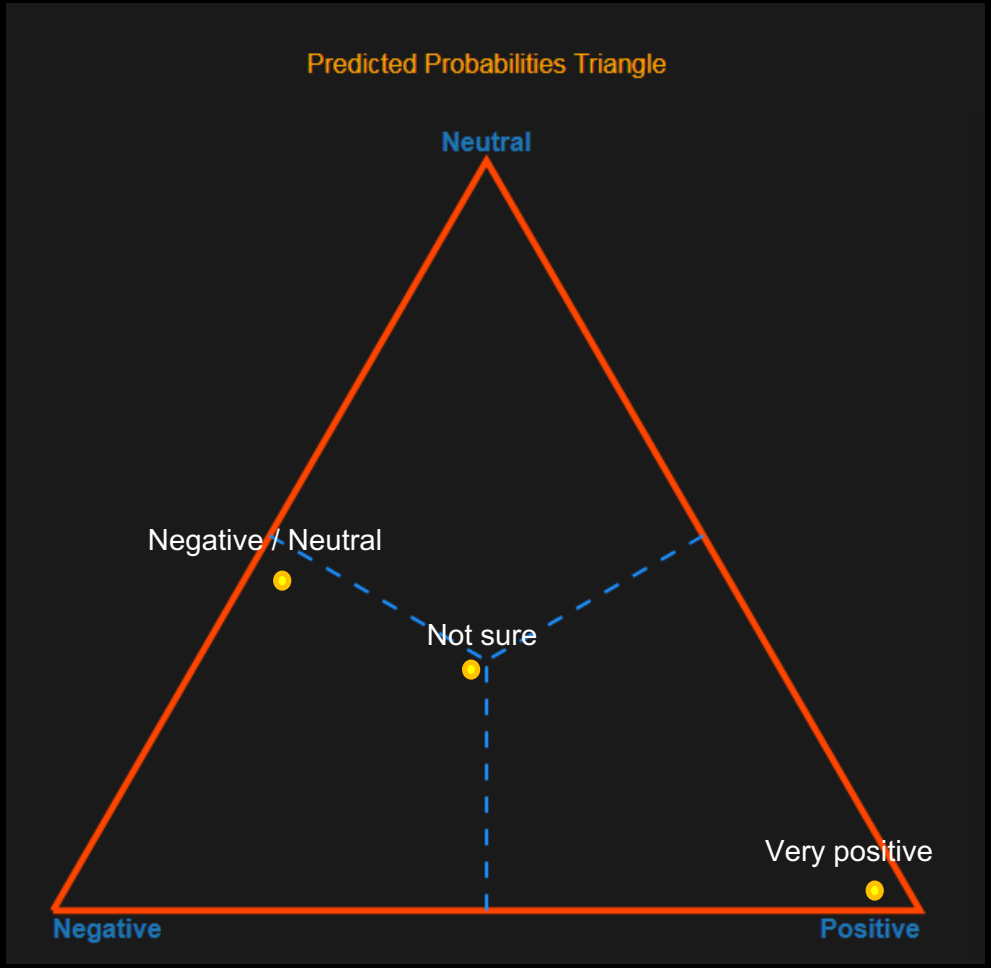
TechAtBloomberg.com

Bloomberg

Triangle Visualization

- Model returns 3 probabilities (which sum to 1)

- How can we visualize these 3 numbers?
  - Points inside an equilateral triangle



TechAtBloomberg.com

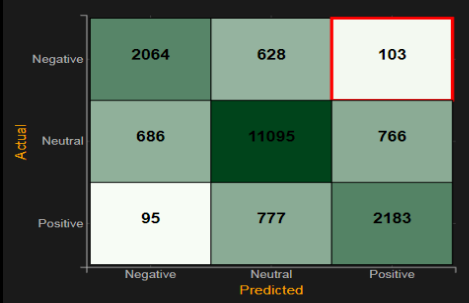
Bloomberg

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)

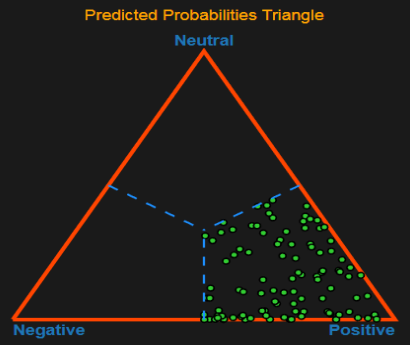
Analyze Misclassifications

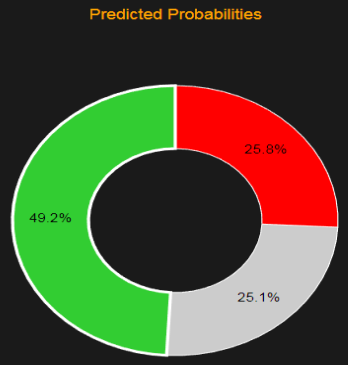


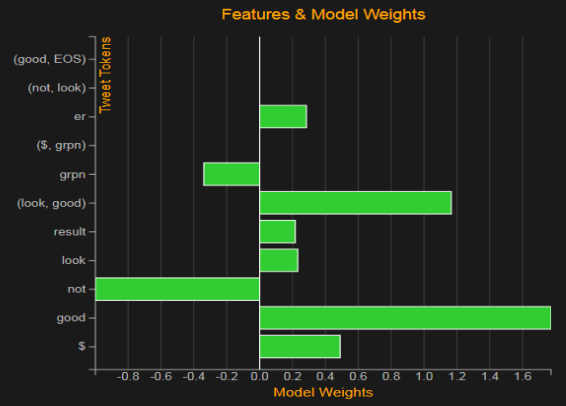
Actual: Negative, Predicted: Positive (103 tweets)

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 IOSP Stock year to date has changed -2.73% percent. +11.37% in the last 30 days. $IOSP 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 highligt 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



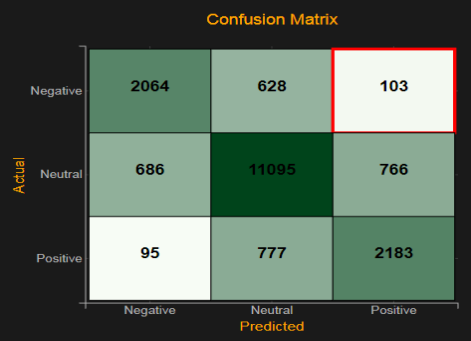




TechAtBloomberg.com

Bloomberg

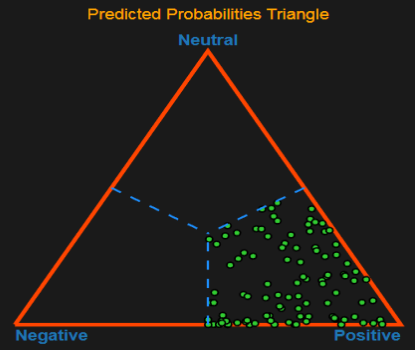
Analyze Misclassifications

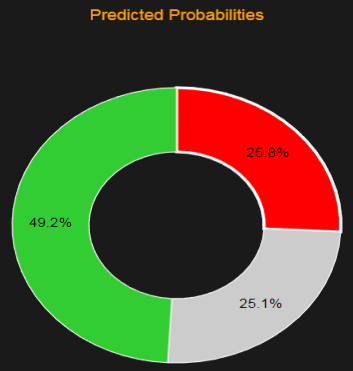


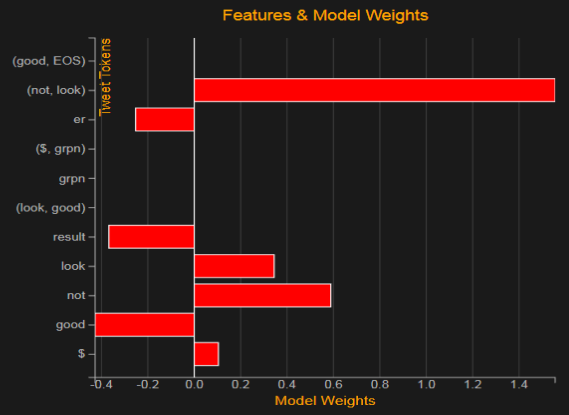
Actual: Negative, Predicted: Positive (103 tweets)

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 IOSP Stock year to date has changed -2.73% percent. +11.37% in the last 30 days. $IOSP 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 highligt 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



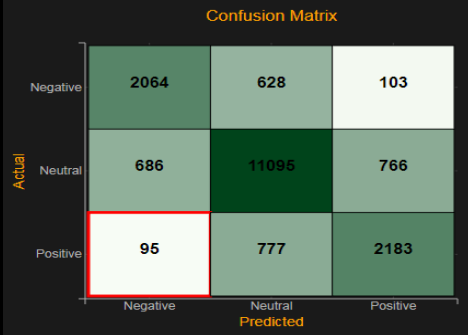




TechAtBloomberg.com

Bloomberg

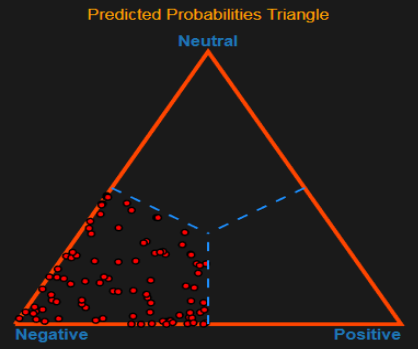
Analyze Misclassifications

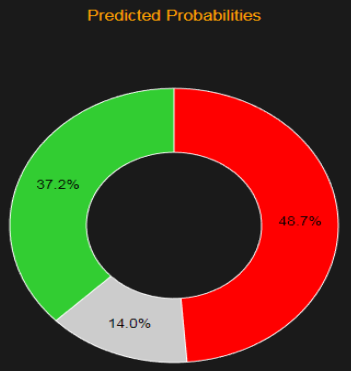


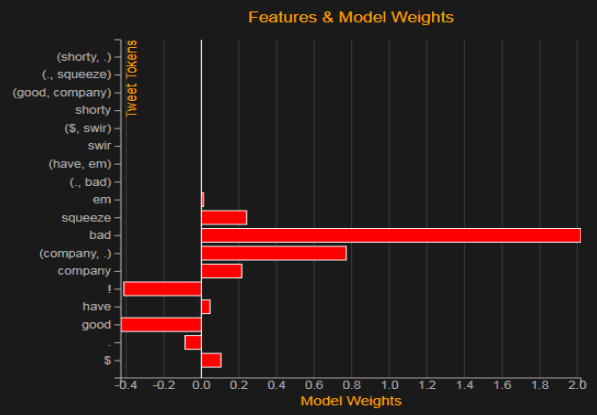
Actual: Positive, Predicted: Negative (95 tweets)

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 shorties. 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 ift.tt/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 shorties. Squeeze 'em! http://stks.co/q361z



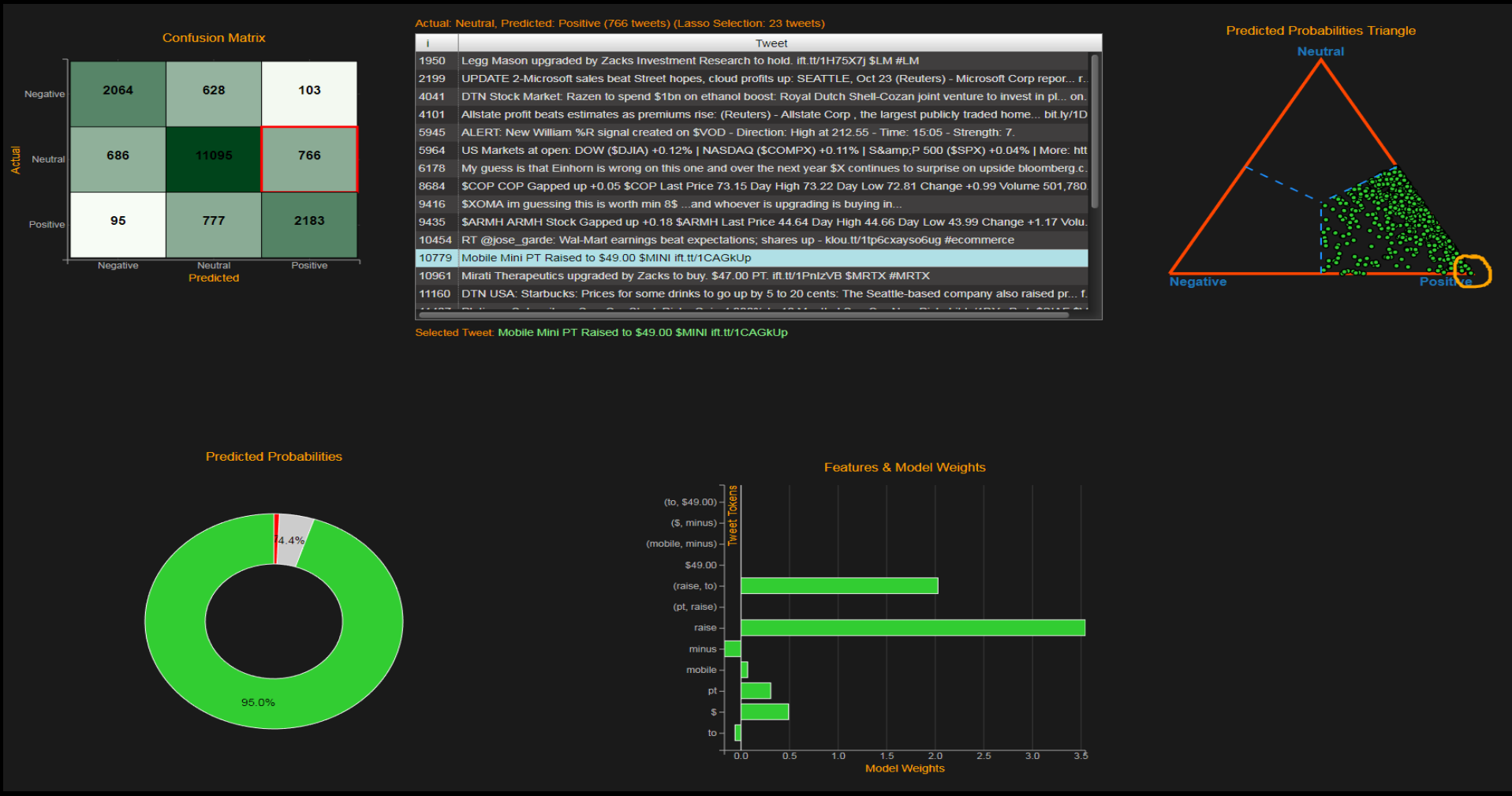




TechAtBloomberg.com

Bloomberg

Use Lasso To Find Data Issues



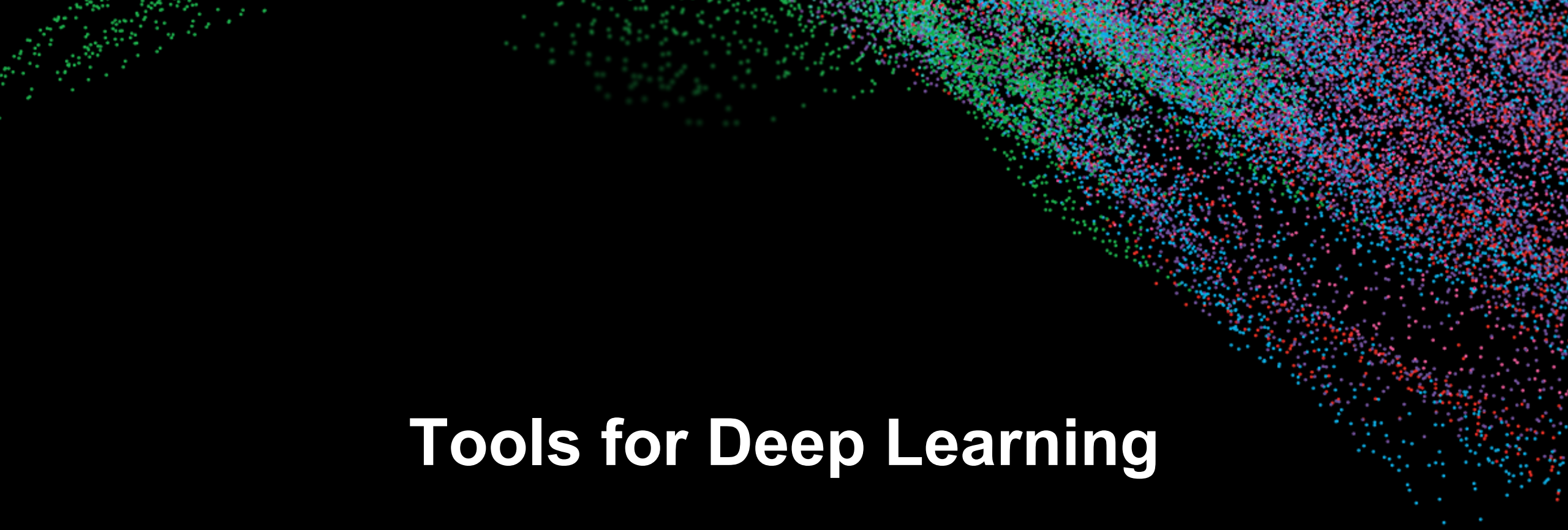
TechAtBloomberg.com

Bloomberg

Use Lasso To Find Data Issues

TechAtBloomberg.com

Bloomberg



Tools for Deep Learning

TechAtBloomberg.com

Bloomberg

Deep Learning Tools

Graphical Wizard

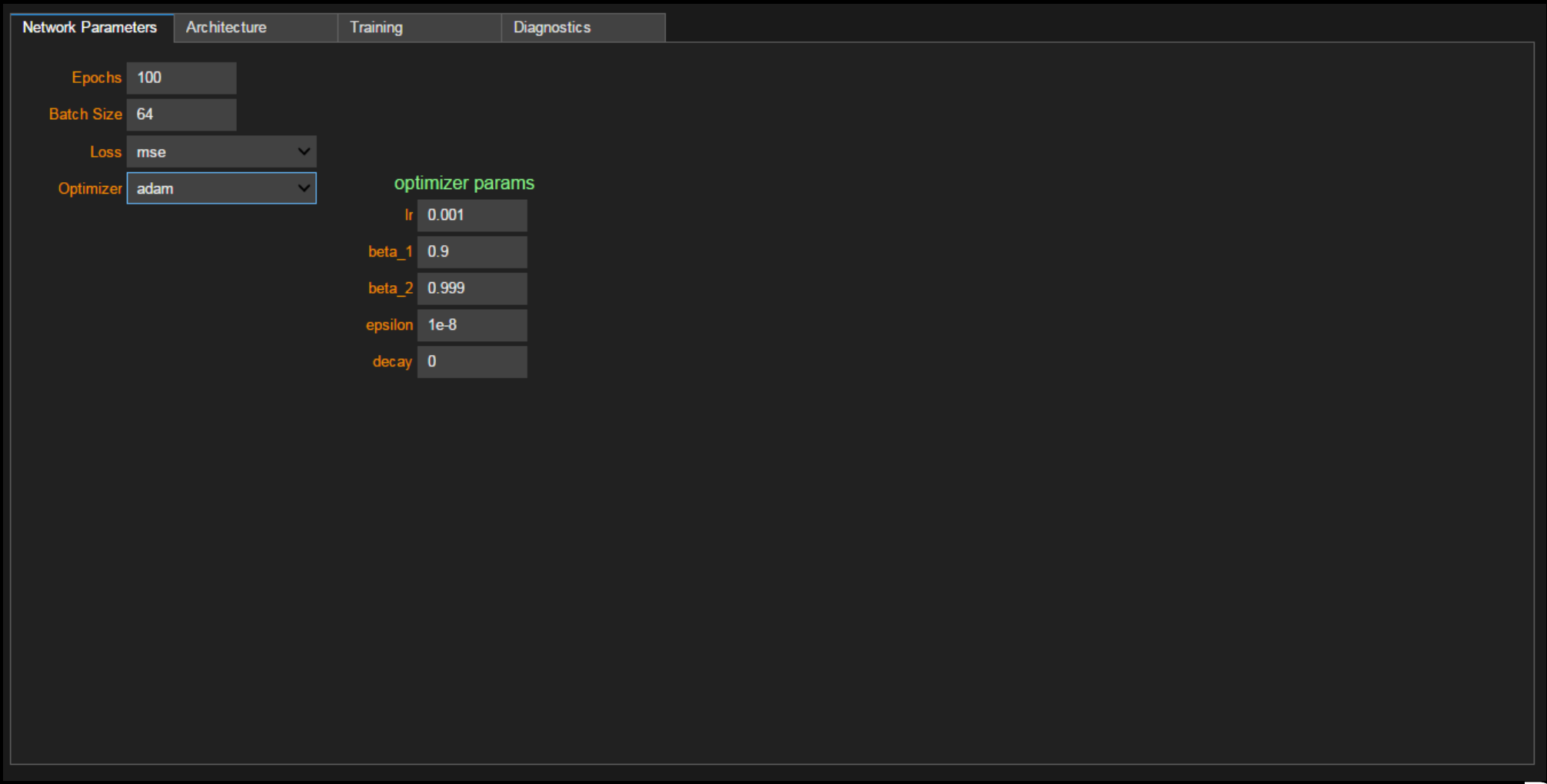
— Select network parameters

— Build network architecture

— Training plots
- Real time loss/accuracy curves
- Distribution of weights/biases/activations

— Diagnostic plots
- Residual vs. Predicted Values
- Confusion Matrix

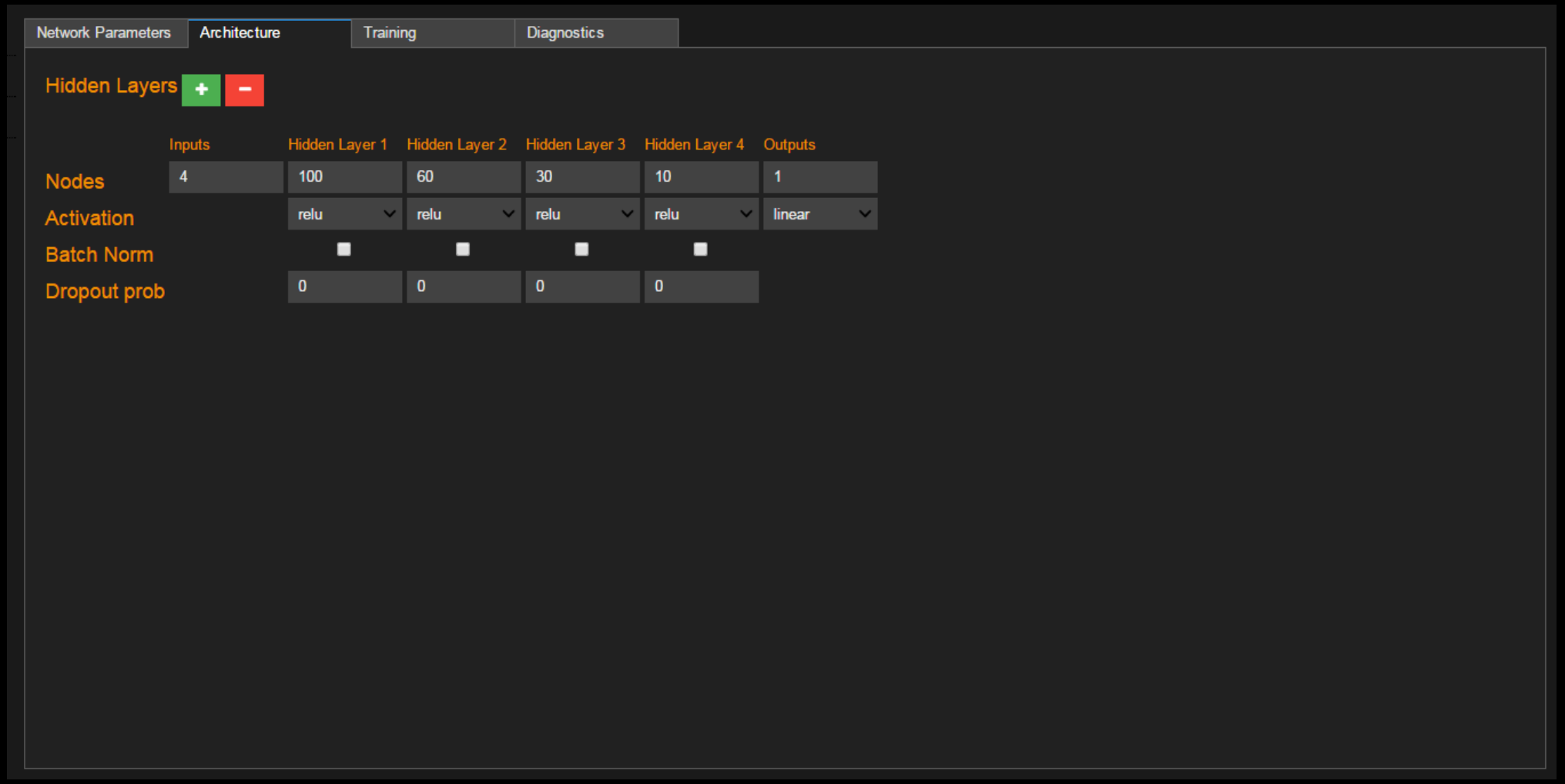
Network Parameters



TechAtBloomberg.com

Bloomberg

Network Architecture



TechAtBloomberg.com

Bloomberg

Loss/Accuracy Curves

TechAtBloomberg.com

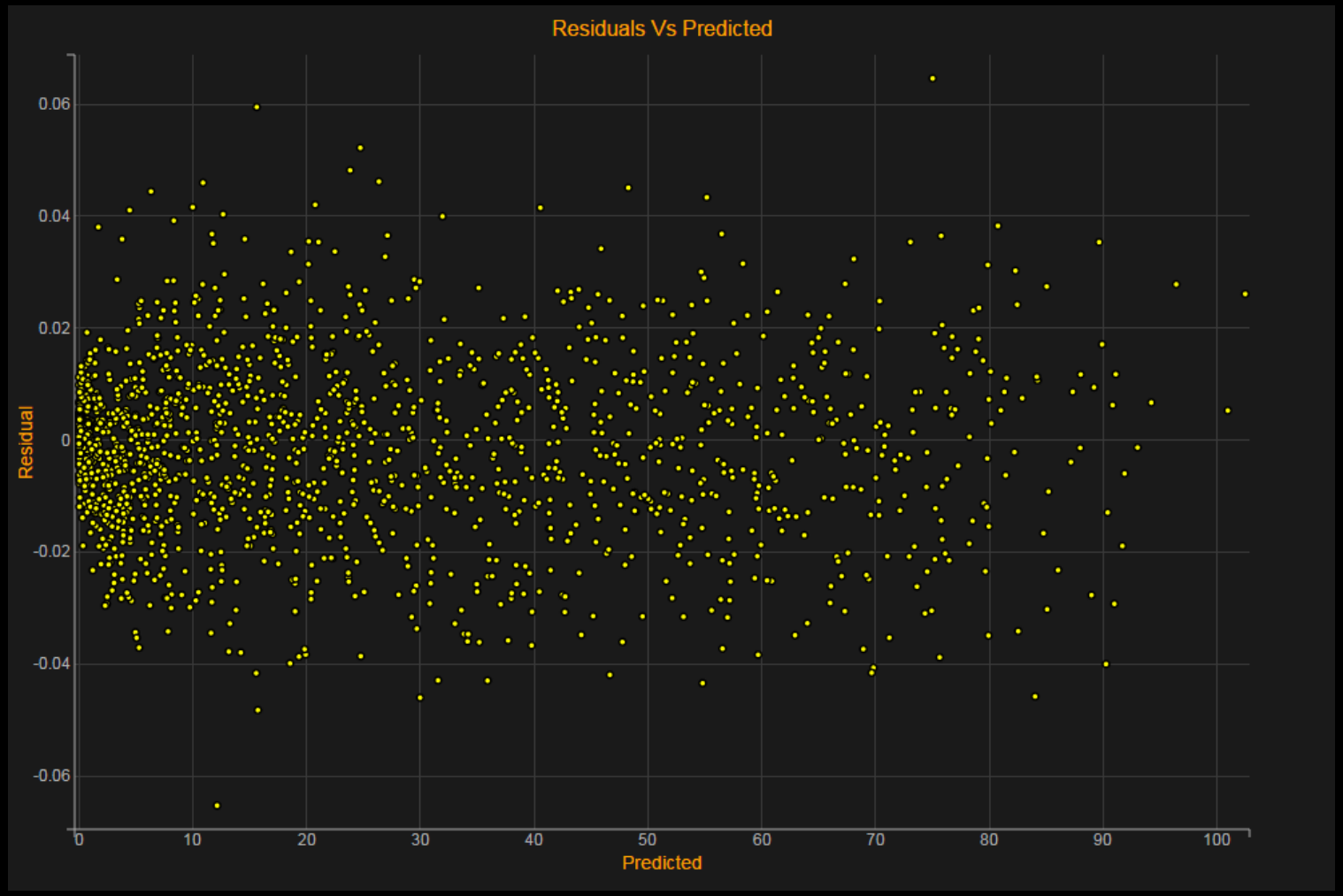
Bloomberg

Distributions of Weights/Biases/Activations

TechAtBloomberg.com

Bloomberg

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

- Tech At Bloomberg blog: www.TechAtBloomberg.com