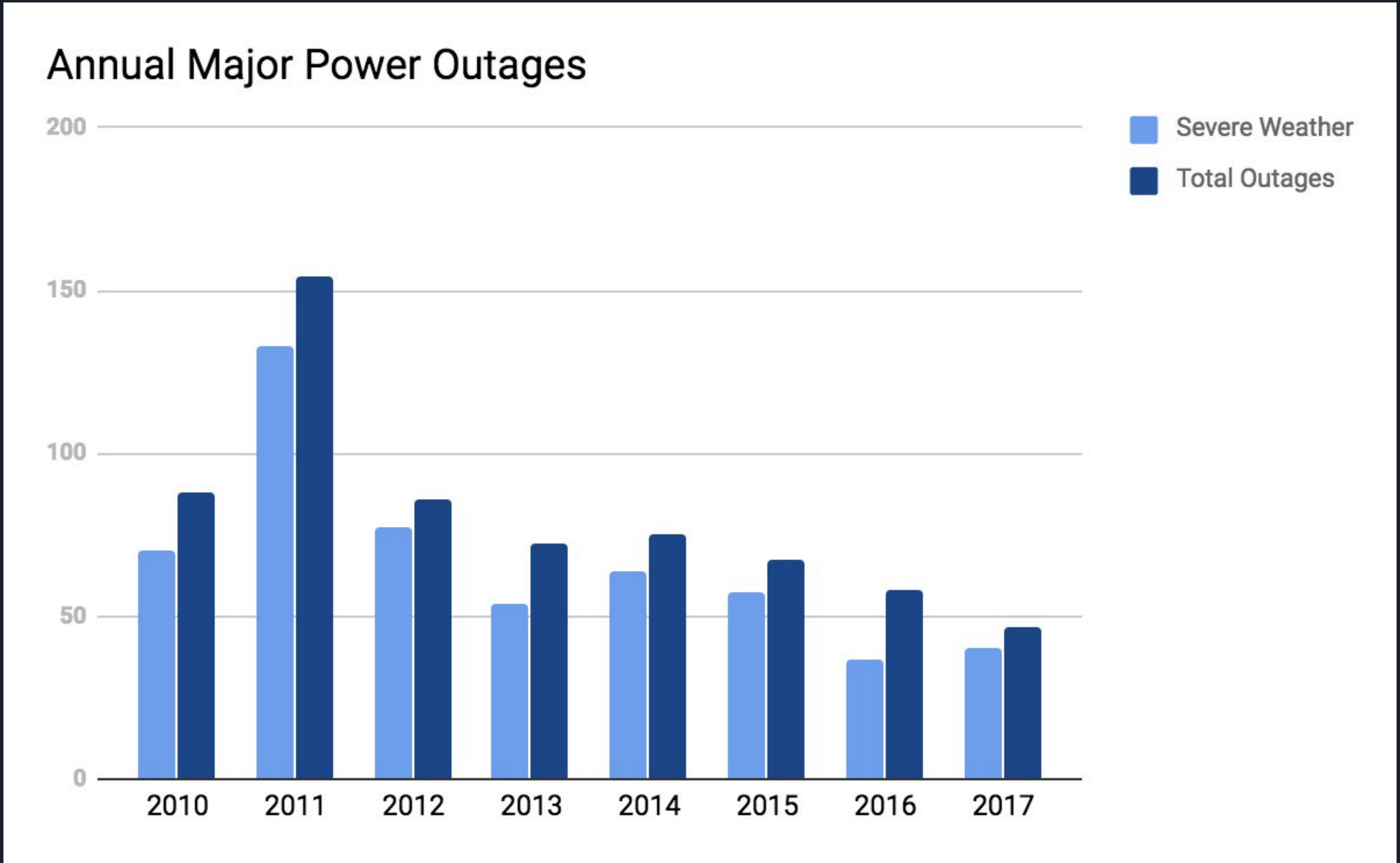
Power Outage Predictor

Predicting the possibility of a power outage in the US based on the weather forecast

- Shruthi Sivasubramanian
- Siva Tejaswi Irkm
- Abhay Sarda

Introduction

- Previous literature[1,2] has investigated in depth:
    - The cost of a power outage to providers and consumers, and
    - The probability of occurrence of such an event from the energy providers' perspective
- We would like to investigate the probability of occurrence of such an event from the consumer's side
- We will focus on weather related power outages


Annual Major Power Outages
Severe Weather
Total Outages
200
150
100
50
0
2010
2011
2012
2013
2014
2015
2016
2017

Literature Review

- Davidson et al. [3] analyzed power outages caused by hurricanes and found a statistically significant relationship between maximum wind speed squared and the probability of outages.
- Liu et al. [4] extended the model by developing generalized linear models. They found that the most important variables were:
    - Maximum wind gust speed
    - Precipitation
    - Temperature
    - Dew Point

Datasets used

- Historical weather data records from the National Oceanic and Atmospheric Administration (NOAA), Climate Data Online.[5]
- Annual power outage summaries from the Department of Energy[6]
- Severe weather inventory[7]
- Storm Prediction Center Reports[8]
- Daily Global Surface Summary Data[9]

Data Analysis

- As a proof of concept, we picked the states which have had the most number of power outages in the past 8 years for our analysis.
- The states which have been affected with the most power outages are Texas and Michigan.
- From those two states, we again picked cities which had the most power outages. Houston for Texas and Ann Arbor from Michigan.
- We combined the weather dataset with the power outage reports, by marking the days during which there was a power outage.

Data Analysis

- The weather data that gave us the best results was the Daily Summaries Dataset from NOAA.
- It had the following fields of data for each day in the time period:
  - Max. 5 second wind speed, Max. 2 minute wind speed, Max Avg. wind speed
  - Precipitation
  - Tmax and Tmin (Together giving the temperature variance in a day)
  - Indicators for Hail, Fog/Ice, Heavy/Freezing Fog and Thunder
- We added three columns for the wind speed squared for each type, and a new column for delta T in a day, based on past literature.

Data Analysis (Cont.)

- The data was scaled using the MinMaxScaler module from SKLearn, to a range(0,1).
- We have several weather variables, but only a few of them will be useful for predictions. To narrow down the list, we first eliminate collinear variables, and then find the most significant features based on their predictive power.
- Remove multicollinearity
  - Using the Variance Inflation Factor, eliminate variables with VIF above a certain threshold
  - This eliminated the linear wind speed terms and the maximum and minimum temperature terms

Data Analysis (Cont.)

- Finding the significant features:
  - We used the SelectKBest module from sklearn.feature_selection
  - The metric for comparison was using chi squared scores, to isolate the main causes for power outages.
  - The selected features after thresholding are:
    - 5 second wind speed squared
    - Avg wind speed squared
    - 2 minute wind speed squared
    - Fog/Ice
    - Heavy Freezing Fog
    - Thunder

Data Analysis (Cont.)

- Our data set is highly imbalanced, with almost 92% of the data having no power outages.
- To train any classifier, you need a significant portion of the data to be of both classes.
- To fix this problem, we tested the following techniques:
  - Undersampling the majority class : K means cluster centroids
  - Oversampling the minority class : Synthetic Minority Oversampling technique
- The former gave better and more consistent results for the data.

Machine Learning Algorithms

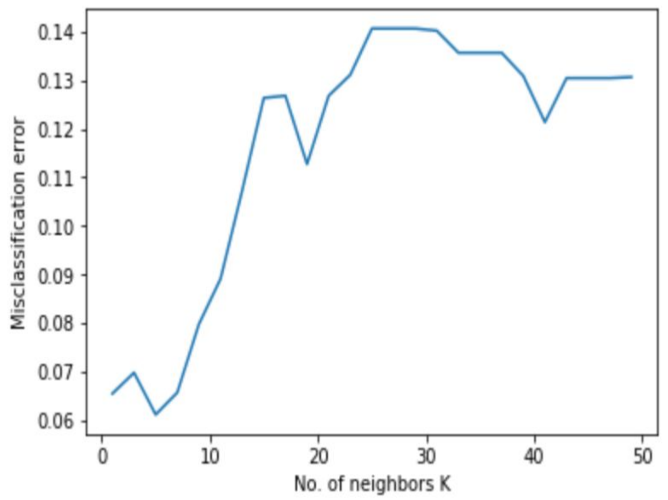
- Since predicting power outages can be modelled as a classification problem, we tried different classification techniques.
- All models tested were given the six features previously identified.
- The data was split into training and testing data in the ratio 2:1
- Some of the ones which are not mentioned in detail because of poor performance (<90% accuracy) are:
    - Naive Bayes
    - Logistic Regression
    - Support Vector Machines
    - Decision Trees

K Nearest Neighbors

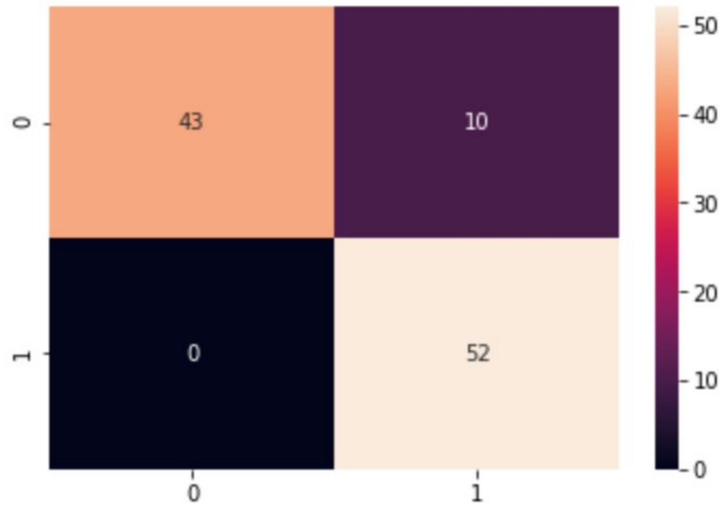
- Over time, similar weather conditions should have similar probabilities of having power outages. This seems like an apt problem for K Nearest Neighbors.
- We tested out both the weighted(distance) and uniform classifiers.
- To cross-validate our results, and determine the ideal value of k, we used k-fold cross validation, with k=10.
- Results:
    - The weighted classifier performed better than the unweighted classifier.
    - The ideal value of k in nearest neighbor search is 5

K Nearest Neighbors (Cont.)

```
Optimal K value is :  5
Misclassification error at that value is  0.0611255411255
```




```
Accuracy:  0.904761904762
```

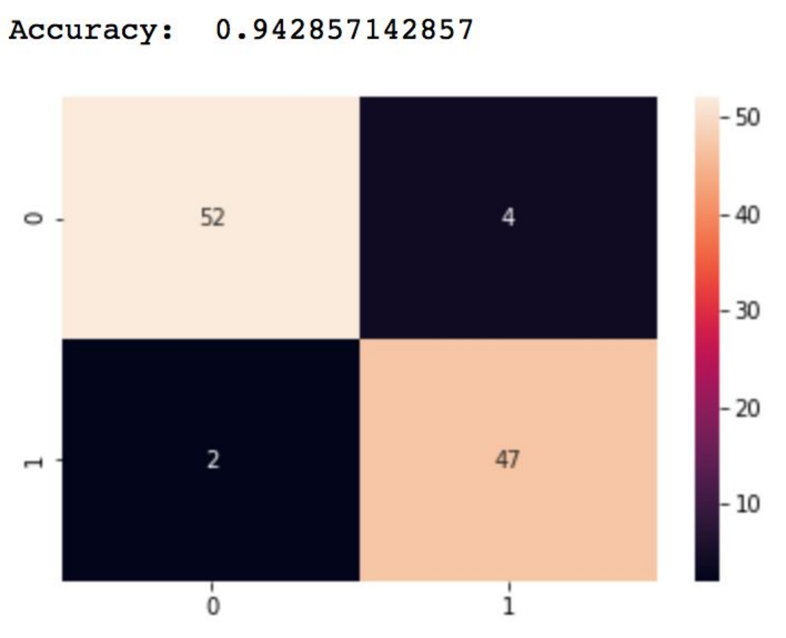


Random Forest Classifiers

- Ensemble methods combine predictions of several base estimators to improve robustness over a single estimator
- We tried both averaging methods and boosting methods for random forest:
  - Averaging methods
    - RandomForest Algorithm
  - Boosting Methods
    - Adaboost
    - Gradient Tree Boosting

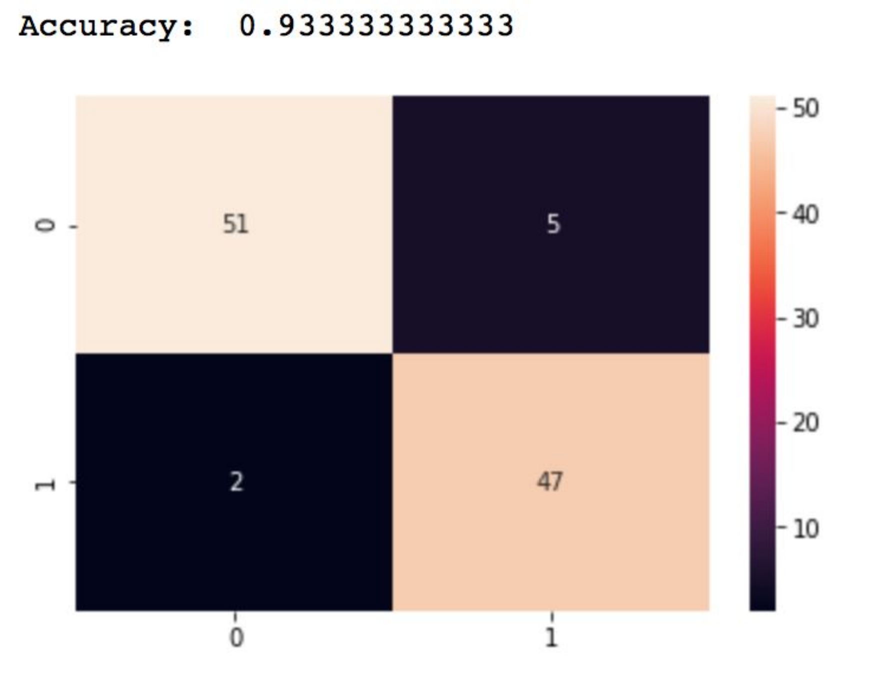
Random Forest Algorithm (Averaging)

- Each tree is built from a sample drawn with replacement.
- No. of estimators: 50
- Max_features: 6
- Accuracy: 94.28%



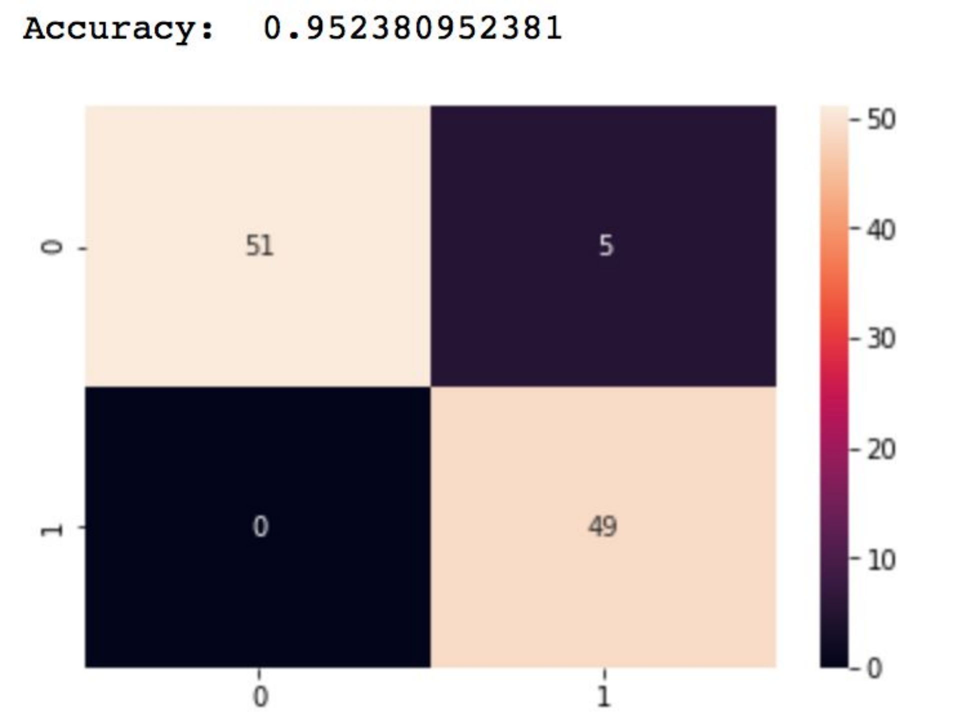
AdaBoost Algorithm

- Fits a sequence of weak learners to repeatedly modified versions of the data.
- Predictions are combined through a weighted majority vote
- No. of estimators : 100
- Accuracy: 93.33%



Gradient Tree Boosting Algorithm

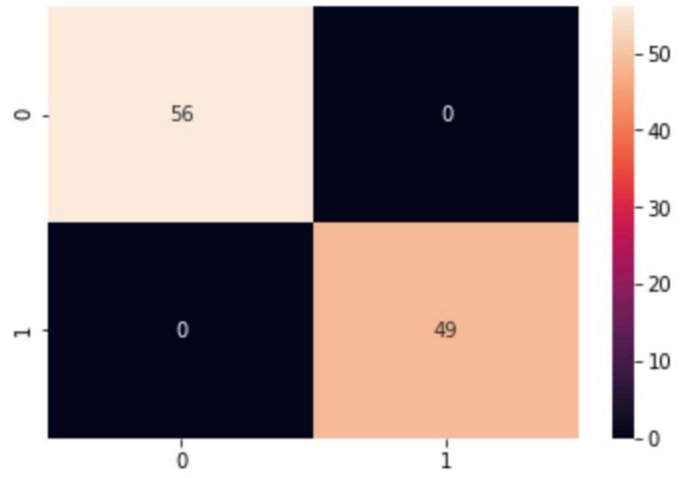
- Robust to outliers in the output space due to robust loss functions
- No. of estimators: 100
- Learning rate : 1.0
- Accuracy: 95.23%



Voting Classifier

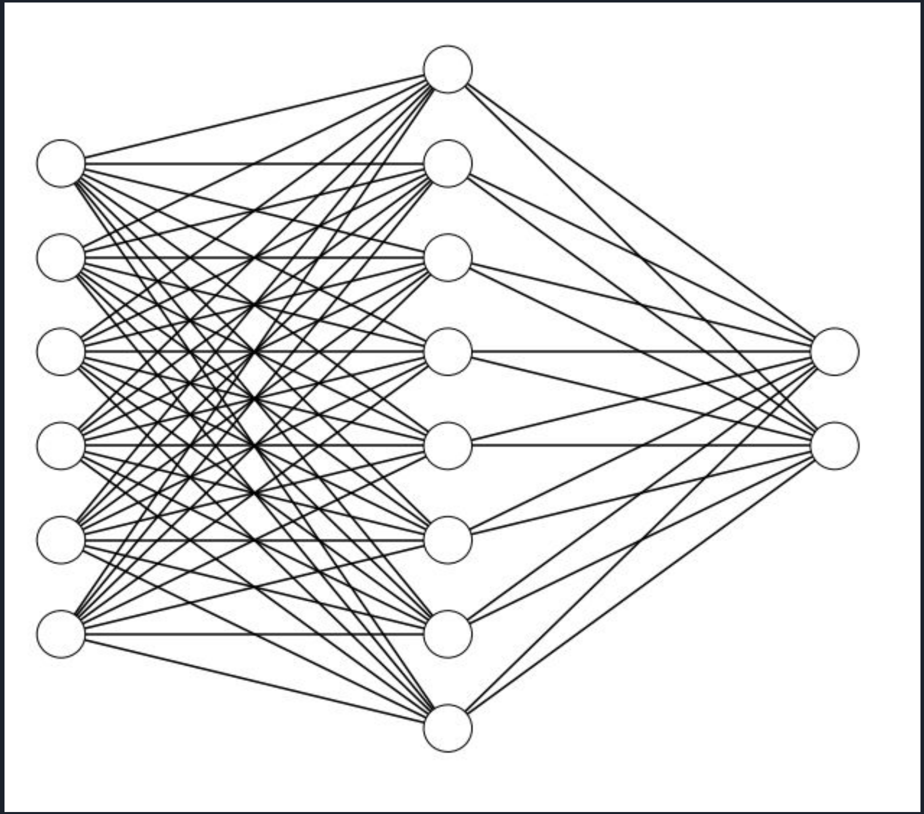
- Classifiers used:
  - K Nearest neighbors
  - Random Forest
  - Gradient Boost
- Voting: Hard
- Accuracy: 95.23%

```
Accuracy for cross validation run:
Accuracy: 0.92 (+/- 0.06) [K Nearest Neighbors]
Accuracy: 0.92 (+/- 0.07) [Random Forest]
Accuracy: 0.94 (+/- 0.04) [Gradient Boost]
Accuracy: 0.94 (+/- 0.05) [Ensemble]
Accuracy for test data
Accuracy:  0.952380952381
```



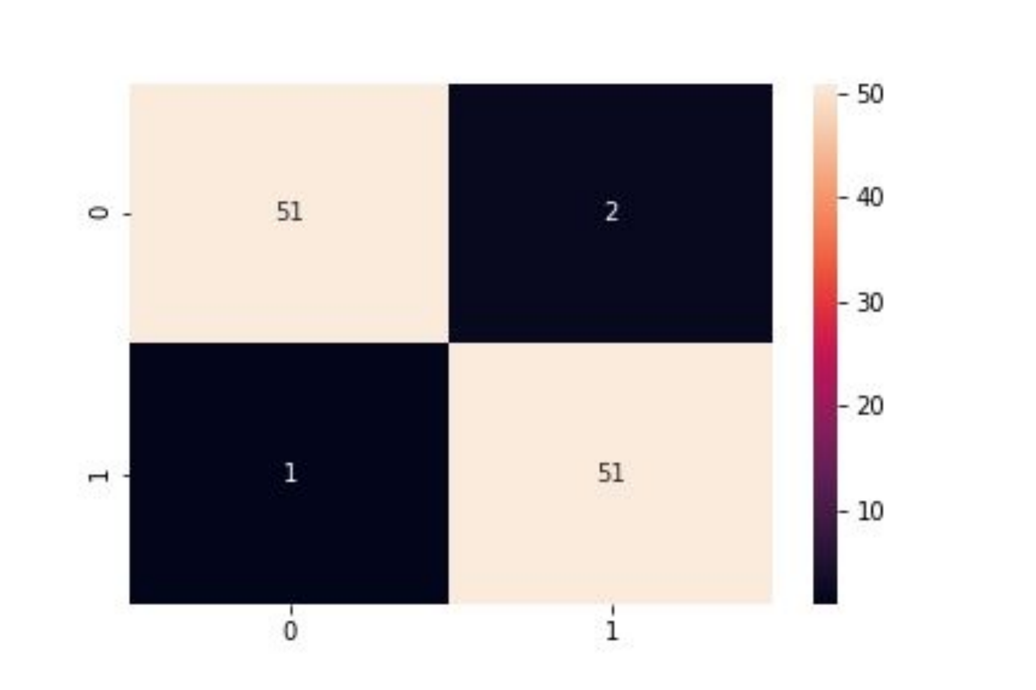
Neural Network Implementation

- Classification problem.
- 1 hidden layer of 8 nodes.
- J(theta) minimization using ‘lbfgs’ solver.



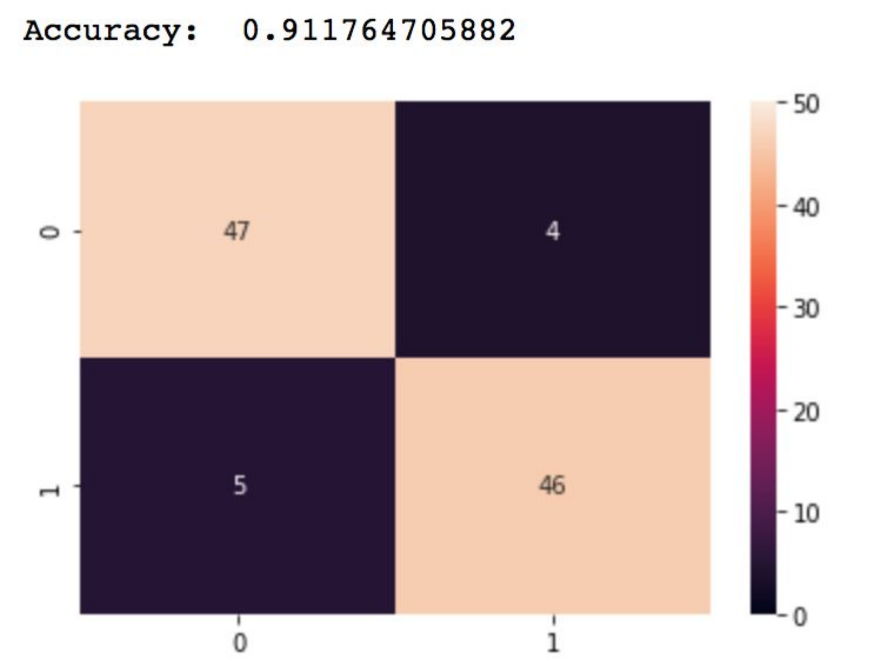
Neural Network Implementation (Cont.)

- Accuracy: 97.14%
- Runs: 1000



Conclusion

- We trained our Voting Classifier on Michigan's weather and tested it on Indiana.
- Indiana is a neighboring state, also struck by many of the same storms.



Future Work

- Our model predicts the possibility of a power outage with high certainty.
- When extended to the entire dataset(without undersampling), it labels normal days as power outages because of bias.
- Ensemble predicts:
  - Actual Power Outages: 100%
  - Normal days: 63%
- We would like to train our model sufficiently for it to predict both with higher accuracy.

Thank you!

References

1. Z. Huang, D. Rosowsky, and P. Sparks, “Hurricane simulation techniques for the evaluation of wind-speeds and expected insurance losses,” J. Wind Eng. Ind. Aerodyn., vol. 89, no. 7, pp. 605–617, 2001.
2. B. J. Cerruti and S. G. Decker, “A statistical forecast model of weather related damage to a major electric utility,” Appl Meteor Clim., no. 51, pp. 191–204, 2012.
3. H. Liu, R. A. Davidson, D. V. Rosowsky, and J. R. Stedinger, “Negative binomial regression of electric power outages in hurricanes,” J. Infrastruct. Syst., vol. 11, no. 4, pp. 258–267, 2005.
4. H. Liu, R. A. Davidson, and T. V. Apanasovich, “Statistical forecasting of electric power restoration times in hurricanes and ice storms,” Power Syst. IEEE Trans. On, vol. 22, no. 4, pp. 2270–2279, 2007.

Datasets

5. NOAA Climate Data Online: https://www.ncdc.noaa.gov/cdo-web/
6. Electric Disturbance Events, DOE:
   https://www.oe.netl.doe.gov/OE417_annual_summary.aspx
7. SPC Reports: https://www.kaggle.com/noaa/noaa-spc
8. Storm Prediction Reports: https://www.kaggle.com/jtennis/spctornado
9. Daily Summaries data : https://www.kaggle.com/noaa/gsod