# Automatic Customer Ticket Classification Model Design and Implementation


ELECTRICAL ENGINEERING
UNIVERSITY of WASHINGTON

Project Members: Swetha Kannan, Huy Nyugen, Ruolan Weir, Zehao Hu

Mentors: Xiang Chen, Payman Arabshahi, Sreeram Kannan

TUPL


## MOTIVATION

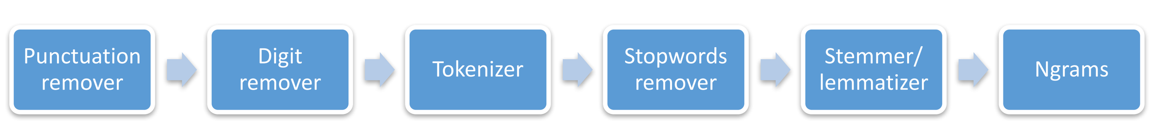
An average telecom network operator receives nearly 300,000 customer complaint tickets in a year. The latency time between complaint reception and resolution is very high, thereby incurring huge operational costs on the network operator side and dissatisfaction on the customer side. This project in collaboration with Tupl Inc. aims to maximize the benefits for the network operators and their customers. Customer complaints can be designated to 9 different issues. Machine learning models process these complaint tickets and classify them into appropriate categories. This model has the potential to automatically process nearly 80% of the tickets without increasing Engineers' workload.

## APPROACH

**Dataset:** 11562 customer complaint tickets.
**Train/ Test split:** 80% training, 10% validation, 10% testing.

**Multinomial Naïve Bayes:** Naive Bayes classifiers, a family of classifiers that are based on the popular Bayes' probability theorem, are known for creating simple yet well performing models, especially in the fields of text/ document classification.

- **Pre-processing:** The raw data is pre-processed by applying several feature transformers. These include punctuation removal, digit removal, tokenization, stop words removal and stemming/ lemmatization.




- **Classification:** Multinomial Naïve Bayes classifier model was used with smoothing.


**Long Short Term Memory Networks (LSTMs):** A special kind of Recurrent Neural Network (RNN) capable of learning long-term dependencies.

- **Pre-processing:** Punctuations and stop words are removed from the raw data.
- **Word-embeddings:** Word-embeddings are obtained using word2vec algorithm.
- **Visualization:** The word vectors are visualized using t-distributed stochastic neighbor embedding (t-SNE) for dimensionality reduction.
- **LSTM Model:** Word vectors are fed into LSTM and trained to obtain a classifier that performs optimally.

**System Integration:**

- **Speech Recognition:** Open source voice recognition software Pypl is used to convert speech to text.
- Front-end/ Back-end: JavaScript and React was used to design the front-end. The back-end framework was developed using Flask.
- **Website Development:** REACT was used to build the website.
- **System Design:** The API is developed using REST API and integrated with Naïve Bayes Classification model.


## SOFTWARE/ PACKAGES USED

Apache Spark, Apache Zeppelin Notebook, Python, Pandas, NLTK, Numpy, Tensorflow, Pypl, React, REST API, JavaScript.

## PERFORMANCE METRIC

Accuracy is the evaluation metric used for the project. The goal is to maximize this value.

Accuracy = #no of cases (predicted labels = actual labels)

## RESULTS AND ANALYSIS I

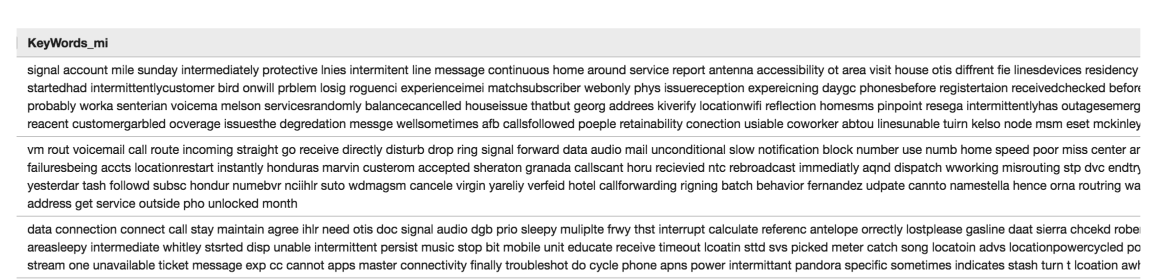
**Pre-processed data:** Several combinations of feature transformers were tested before deciding the final pre-processing feature transformers. Lemmatization instead of stemming improved the performance of the model slightly.

| Punctuation_removed | Digit_removed | Tokenized | Stop_wordsremoved | wrongspellings_removed |
|---|---|---|---|---|
| cx is having prob... | cx is having prob... | [cx, is, having, ... | [cx, problems, sl... | [problems, slow, ... |
| cust keeps losing... | cust keeps losing... | [cust, keeps, los... | [cust, keeps, los... | [keeps, losing, s... |
| Customer opened u... | Customer opened u... | [customer, opened... | [customer, opened... | [customer, opened... |
| weak signal KENN... | weak signal KENN... | [weak, signal, ke... | [weak, signal, ke... | [weak, signal, we... |
| Incoming calls o... | Incoming calls o... | [incoming, calls,... | [incoming, calls,... | [incoming, calls,... |
| Created 123 Stree... | Created StreetNa... | [created, streetn... | [created, streetn... | [created, nm, che... |
| has extreme slown... | has extreme slown... | [has, extreme, sl... | [extreme, slownes... | [extreme, slownes... |
| Not able to load ... | Not able to load ... | [not, able, to, l... | [able, load, web,... | [able, load, web,... |
| keeps on loosing ... | keeps on loosing ... | [keeps, on, loosi... | [keeps, loosing, ... | [keeps, loosing, ... |
| Customer is not a... | Customer is not a... | [customer, is, no... | [customer, able, ... | [customer, able, ... |
| cannot call numb1... | cannot call numb ... | [cannot, call, nu... | [cannot, call, nu... | [cannot, call, nu... |
| Cx can not send a... | Cx can not send a... | [cx, can, not, se... | [cx, send, receiv... | [send, receive, p... |
| voicemail giving ... | voicemail giving ... | [voicemail, givin... | [voicemail, givin... | [giving, error, n... |
| Cannot receive a ... | Cannot receive a ... | [cannot, receive,... | [cannot, receive,... | [cannot, receive,... |

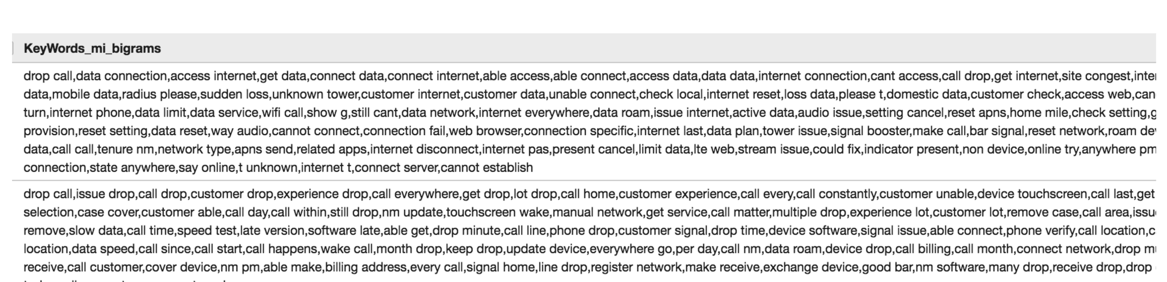
**Keywords:** Term frequencies and mutual information was used to obtain keywords from the data.

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

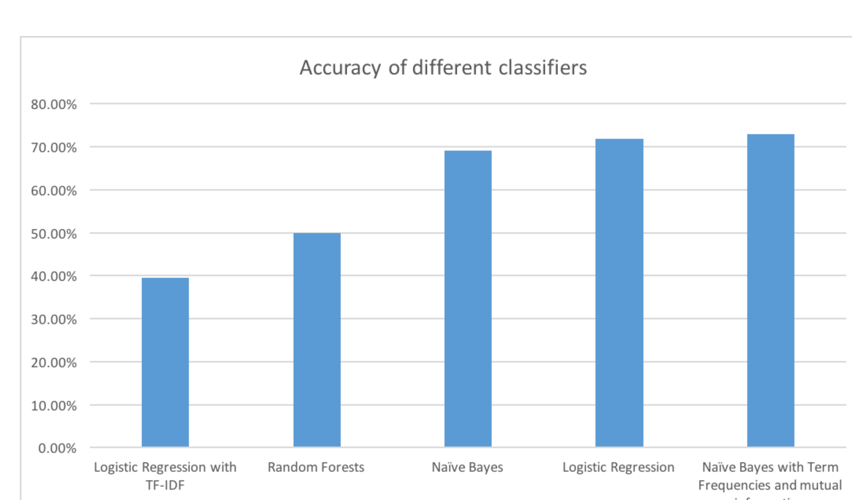
Unigrams using mutual Information:



Bigrams using Mutual Information:

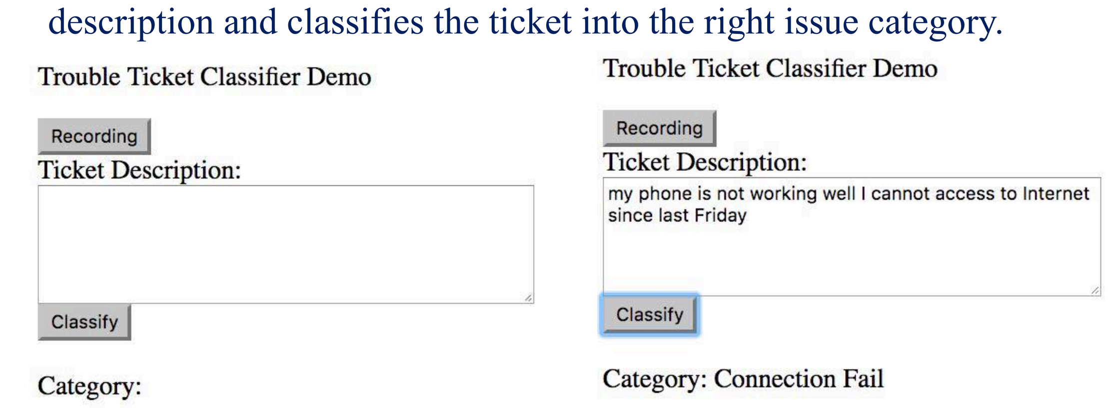


**Comparison of Classifiers:** Naïve Bayes with Term frequencies and mutual information had the best accuracy compared to all other models.

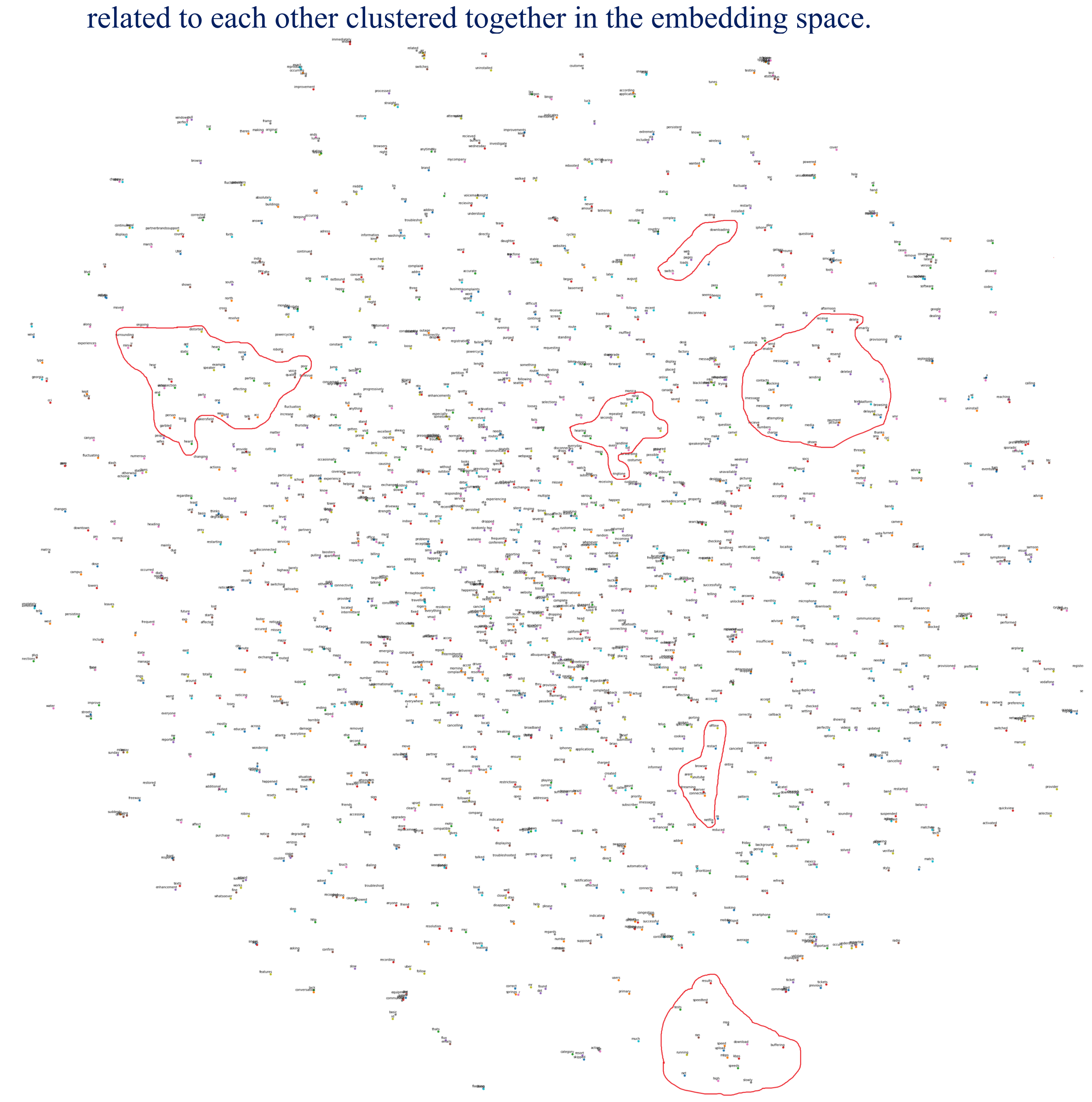



## RESULTS AND ANALYSIS II

**Integrated System:** The interactive website takes in an audio recording/ description and classifies the ticket into the right issue category.



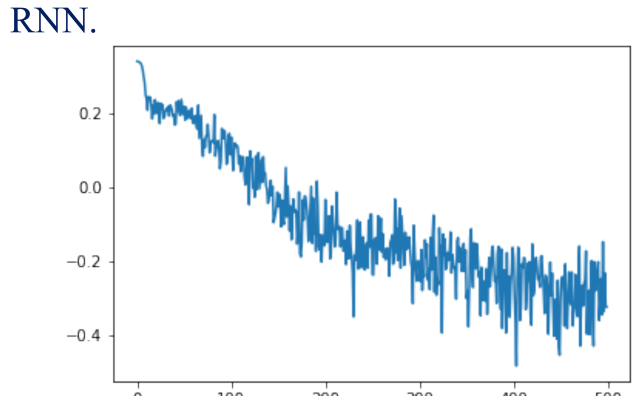

**Embedding Visualization:** The word embeddings were visualized using t-SNE dimensionality reduction technique. It was found that several words related to each other clustered together in the embedding space.



**RNN Pre-processing:** All the tickets were padded with starting and ending tokens until all tickets were of equal length. The original word lengths were also computed and fed to the RNN model.

| splitWords | wordLengths | wordsPadded | wordsId |
|---|---|---|---|
| [customer, issue, able, get, one, bar, ph, home] | 8 | [<s>, customer, issue, able, get, one, bar, ph... | [872, 109, 790, 176, 69, 761, 553, 1768, 3609,... |
| [states, past, month, 03, 5am, experienced, fl... | 15 | [<s>, states, past, month, 03, 5am, experience... | [872, 3110, 1138, 330, 738, 2103, 1547, 888, 2... |
| [make, receive, calls, signal, bars, verify, n... | 8 | [<s>, make, receive, calls, signal, bars, veri... | [872, 2156, 2718, 1265, 2554, 2993, 716, 2153,... |
| [customer, keeps, dropping, calls, alot, even,... | 13 | [<s>, customer, keeps, dropping, calls, alot, ... | [872, 109, 3259, 3414, 1265, 3543, 1137, 2839,... |
| [cx, able, make, rcv, calls, going, calls, dia... | 59 | [<s>, cx, able, make, rcv, calls, going, calls... | [872, 589, 176, 2156, 3165, 1265, 2535, 1265, ... |

**RNN Training costs:** Cross entropy was the loss measure used for the RNN.



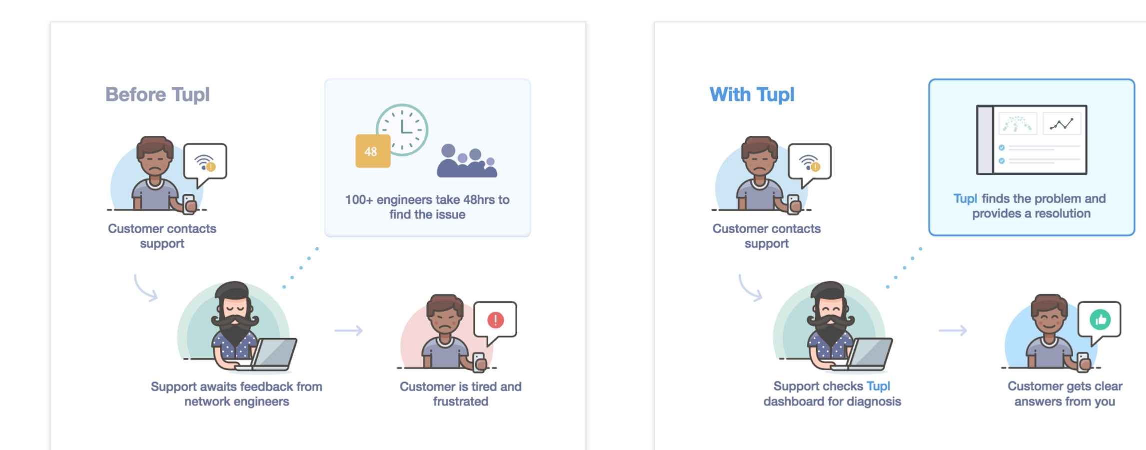

$$H_{y'}(y) = -\sum_i y'_i \log(y_i)$$

## DISCUSSION

**Classifier Accuracy:**
Naïve Bayes Classifier – 75% on test data
LSTM Model – 65% on test data

Currently, the Naïve Bayes classifier outperforms the LSTM Model. This is partly because deep learning requires a lot of tuning and careful selection of hyper-parameters. These include embedding layer size, number of embedding layers, learning rate, drop out, number of epochs etc.

## POTENTIAL IMPACT OF THE PROJECT




- On a scaled level, automation of customer care complaints can save up to 25% of Engineers' workload.
- 80% of complaint ticket resolutions can be automated.
- 10x improvement in responding to customer queries.
- Reduced operational costs for telecom network operators.
- Improved work efficiency for customer support team.
- Improved customer care experience.

## CONCLUSION/ FUTURE WORK

Intelligent Process Automation is a very promising idea that leverages machine learning and big data technologies to offer solutions to various problems. It has a vast potential in the telecom network operations side. The long term goal of this project is to deliver a machine learning classifier with optimal accuracy. The accuracy of the RNN model has to be improved since deep learning is proven to be an interesting direction for several machine learning projects.

## ACKNOWLEDGEMENT


We gratefully acknowledge the Department of Electrical Engineering, Industry capstone course staff and Tupl Inc. for providing the opportunity to work on a real-world problem and provide solutions. We would also specially like to thank our mentors Xiang Chen, Professors Payman Arabshahi and Sreeram Kannan for their guidance.


## REFERENCES


1. https://www.tensorflow.org/tutorials/word2vec
2. http://colah.github.io/posts/2015-08-Understanding-LSTMs/