

# Projects in AI/ML

Author- Mihir N Parikh

## News In-shorts Topic modeling

Description: It tries to create clusters of 60-word news data extracted from New-InShorts app. It will categorize news to several topics based on their content and then label those topics too.

Underlying ML concept: TF-IDF + NMF

Status: Completed, with acceptable accuracy on provided data

Purpose: Academic project

## Object recognition in Images

Description: It was an implementation of mobilenet model to develop an end-to-end tool that can recognize known objects inside images/videos/live feeds and label them. Uses COCO dataset.

Underlying ML concept: Custom CNN on tensorflow

Status: Completed, with good accuracy on real-world data

Purpose: Academic Project

## Document Classifier

Description: It will train a model that is able to classify documents in predefined categories.

Underlying ML concept: TF-IDF + SVM

Status: Completed+Demoable, with 90+ accuracy on provided data

Purpose: PoC for a client famous in Legal industry

## Document Clustering

Description: It will train a model that can create clusters of documents in categories. The number of categories is chosen automatically from the data such that it is best classified in each cluster.

Underlying ML concept: Apache Tika

Status: Completed with Demoable UI

Purpose: PoC for client famous in Housing/Mortgage sector sponsored by US govt

## Student In-take prediction for university

Description: It was a SageMaker based use case where data from redshift was used to predict the number of students that would be admitted in different intake seasons and several prediction graphs were plotted on quicksight portal. It will also predict the number of failing students.

Underlying ML concept: Time Series Analysis

Status: Completed and being used, with acceptable accuracy on provided data

Purpose: PoC for a famous University for Nursing in US

### **Question Suggestion System**

Description: It was designed for forum type sites where the system will try to suggest similar questions that were already answered on the site earlier before the user posts his redundant question. Aka- Recommendation system for Questions

Underlying ML concept: NLP + TF-IDF, Word2Vec + Euclidean Distance

Status: Left in the preliminary stage, was performing well on small sized data.

Purpose: Ideas for Liferay symposium demo

### **Sentiment analysis + Aspect extraction**

Description: Training a model that analyzed reviews/comments of users and tagged the sentiments present on the overall comment as well as sentiments on parts of comments that rate the different aspects of domain-specific items.

Underlying ML concept: Word2Vec + TF-IDF + DL

Status: Left in the preliminary stage, was performing satisfactorily on sample data.

Purpose: Ideas for Liferay symposium demo

### **Stock Market Price Prediction**

Description: It tries to train a model that will predict the open price for some stock for a given day based on historical data.

Underlying ML concept: Deep Learning

Status: Completed, with satisfactory accuracy on provided data

Purpose: Hobby

### **Voice ChatBot for any Sales team**

Description: End-to-end chatbot which can be interacted via text and voice. Entertains questions seeking sales figure for a particular company in a particular year out of a live database.

Underlying ML concept: Custom DL based NER

Status: Completed and Demoable, with good accuracy on normally asked questions

Purpose: PoC for multiple clients famous in finance, consumer products industries

### **Recommendation system for articles**

Description: It will generate recommendations (of related articles) for site users based on their previous interactions within the website. For now, article viewing is considered an interaction.

Underlying ML concept: Collaborative filtering

Status: Completed and Demoable on LifeRay, with good accuracy on generate data

Purpose: Demo at Liferay symposium

### **House Price Prediction**

Description: A use-case to portray our sagemaker capabilities which focused on predicting the prices at which the house may be sold based on various parameters of that house.

Underlying ML concept: XgBoost

Status: Completed, with acceptable accuracy on kaggle dataset

Purpose: PoC for a client in Indian govt banking sector

## **Speech API**

Description: A REST API for doing machine translation tasks on human-like voice. It is capable of doing Speech-to-text as well as Text-to-speech for common English accent voices. It leverages Mozilla's DeepSpeech, TTS and pre-trained models available in the community.

Underlying concept: Flask + DeepSpeech + TTS + NER

Status: Completed, with acceptable results on American accented voices.

Purpose: Dependency for PoC of Voice-ChatBot

## **Smart Insights**

Description: This is a system that throws textual crux out of organized data that focuses on interesting insights of data, which might be not directly visible from complex dashboards. It has a feedback system to make generated insights more actionable.

Underlying ML concept: NLG + Correlation + Clustering

Status: Completed, with decisive insights being generated on HR dataset

Purpose: PoC for a client famous in administrative solutions industry