

# Sentiment Analysis on Restaurant Reviews

Ferrin Maria Christina Dsouza  
4SO17CS038

Internal Guide

Mr. Gerald Harry Fernandes

Assistant Professor,

Department of Computer Science  
and Engineering

External Guide

Mr. Guruvansh Singh

Subject Matter Expert

# Contents

- Introduction
- Objectives
- Work carried out
  - About Project
  - Features Implemented
- Software Requirements

- Project Presentation
- Outcomes
  - Technical
  - Non Technical
- Conclusion

# Introduction

- KSI is a Delhi based online skilling and certification company and a Microsoft Authorized Education Partner. The company was incorporated on 15 March in Pune, Maharashtra.
- The company is India's private sector enterprise, with partnership businesses and is an e-learning platform.
- Founded in the year 2006 to help students to get trained, certified and undertake internships.
- The company makes sure that every approach produced operates effectively and delivers tangible results to the customers.



Mr. Sandeep Jethani

**The CEO**

# Introduction

## Services Offered

- The services provided by them are certification, skill development, placement preparation and training.
- The company provides technology and robotics laboratory that is specifically designed to uncover rationale behind technical wonders for young minds aged 12 + and constructs both small business and corporate e-commerce websites.

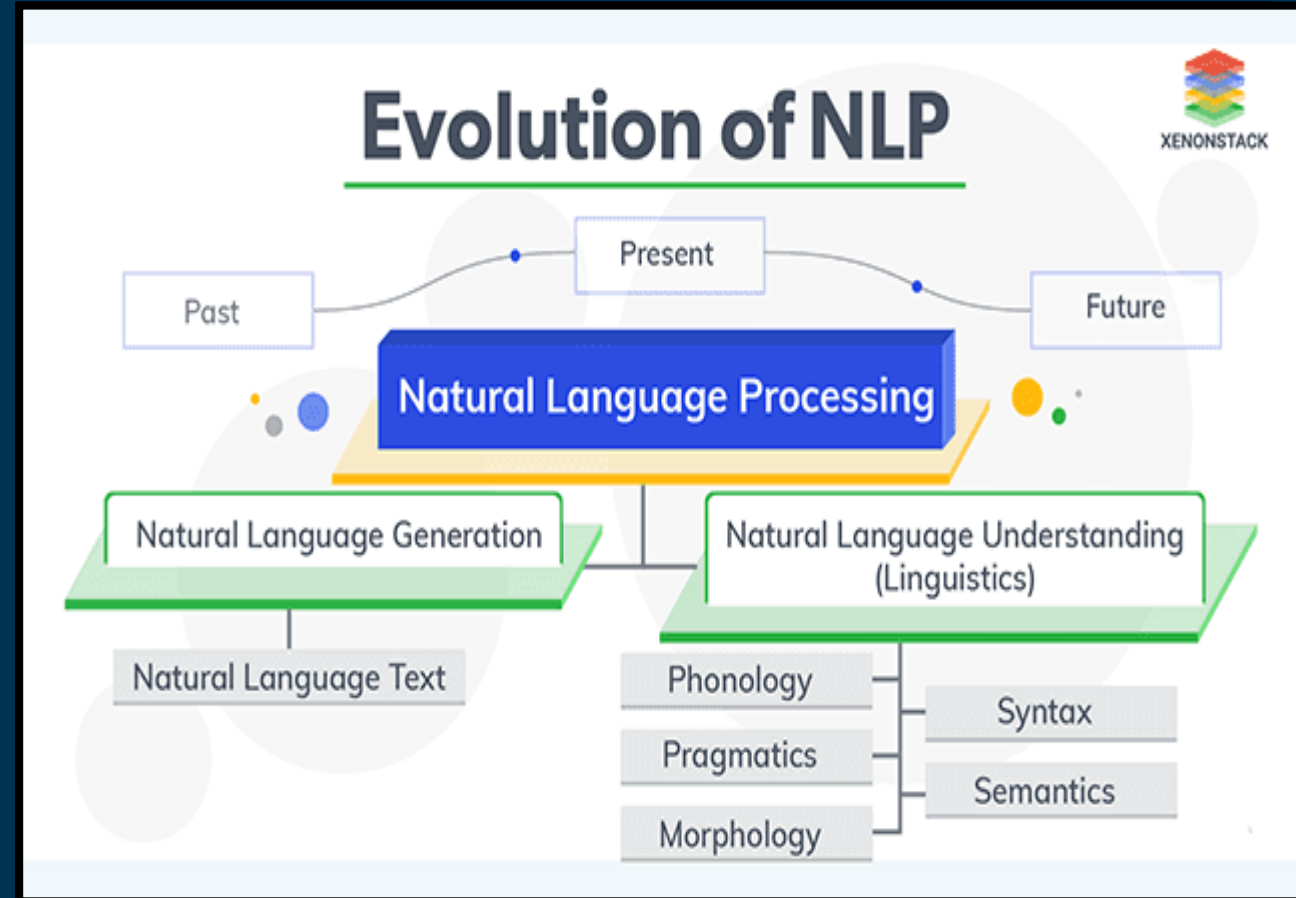


Microsoft in  
Education Global  
Training Partner



# Objectives

- The internship introduces students to machine learning & AI techniques using PYTHON.
- Aim of internship was to gain a clear insight into machine learning applications, exploring and using data sets, supervised vs unsupervised learning techniques etc.



# Work Carried Out

## Project Title: Sentiment Analysis of Restaurant Reviews

- The purpose of this analysis is to build a prediction model to predict whether a review on restaurant is positive or negative.
- This is done by using predictive algorithms such as Multinomial Naïve Bayes, Bernoulli Naïve Bayes and Logical Regression.
- Use model with highest accuracy.
- Build a model to predict.

# Work Carried Out

## - Features Implemented

- Built a prediction model to predict whether a review on restaurant is positive or negative by doing the following steps:
  - Importing a dataset,
  - pre-processing dataset,
  - Vectorization,
  - Training and Classification,
  - Analysis and Conclusion

# Work Carried Out

## - Features Implemented

- Dataset has 1000 rows and 8 columns.
- We have to predict whether the review is “Positive” or “Negative”
- Porter Stemmer method has been used for Stemming.
- Applied many different algorithms LSTM, Bi-directional LSTM, Random Forest Classifier, MultinomialNB, SVM and KNN.
- After performing data cleaning, convert “Rating” column which is actually a numerical column, into the column that has 2 labels “Positive” and “Negative”. Rating above 3 is positive and rating below 3 is negative.



# Software Requirements

- Platform: flask
- Dataset: Kaggle
- Programming Language: Python, HTML, CSS
- Library: NLTK

The screenshot displays the Kaggle Data Explorer interface for the 'Restaurant\_Reviews.tsv' dataset (59.89 KB). The interface includes a search bar, navigation tabs (Data, Tasks, Code (7), Discussion, Activity, Metadata), and buttons for 'Download (60 KB)' and 'New Notebook'. The dataset is categorized with a 'Usability 1.8' score and 'No tags yet'. The 'Data Explorer' section shows a table with two columns: 'Review' and '# Liked'. The table has 977 unique values in the 'Review' column and 705 total values in the '# Liked' column. The table content is as follows:

Review	# Liked
Wow... Loved this place.	1
Crust is not good.	0
Not tasty and the texture was just nasty.	0

At the bottom of the interface, there is a cookie consent message: 'We use cookies on Kaggle to deliver our services, analyze web traffic, and improve your experience on the site. By using Kaggle, you agree to our use of cookies.' with 'Got it' and 'Learn more' buttons.

# Project Presentation



Fig 1. Home Page

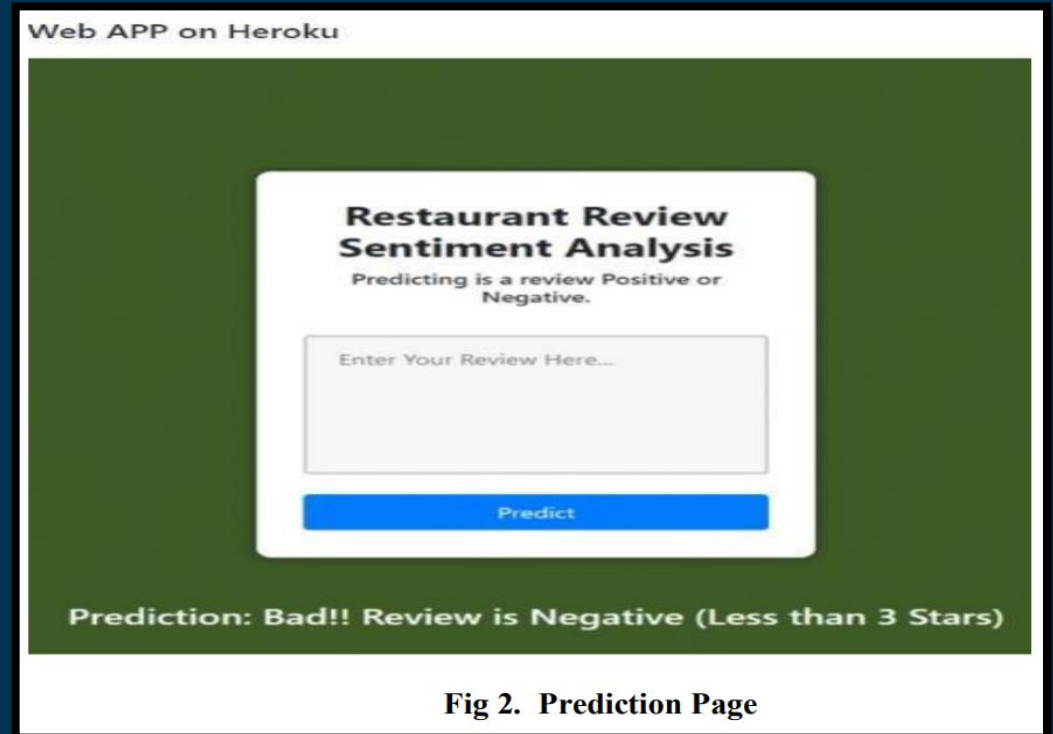


Fig 2. Prediction Page

# Project Presentation

```
In [219]: from tensorflow.keras.layers import Bidirectional
import keras
from tensorflow.keras.layers import Dropout
## Creating model
embedding_vector_features=40
model=tensorflow.keras.Sequential()
model.add(Embedding(voc_size,embedding_vector_features,input_length=sent_length))
model.add(Dropout(0.3))
model.add(Bidirectional(LSTM(150)))
model.add(Dropout(0.3))
model.add(Dense(1,activation='sigmoid'))
model.compile(loss='binary_crossentropy',optimizer='adam',metrics=['accuracy'])

In [220]: model.fit(X_train,y_train,validation_data=(X_test,y_test),epochs=5,batch_size=64)

Train on 7500 samples, validate on 2500 samples
Epoch 1/5
7500/7500 [=====] - 24s 3ms/sample - loss: 0.4506 - accuracy: 0.7997 - val_loss: 0.2628 - val_accuracy: 0.8952
Epoch 2/5
7500/7500 [=====] - 20s 3ms/sample - loss: 0.2333 - accuracy: 0.9065 - val_loss: 0.2500 - val_accuracy: 0.9060
Epoch 3/5
7500/7500 [=====] - 20s 3ms/sample - loss: 0.1949 - accuracy: 0.9281 - val_loss: 0.2491 - val_accuracy: 0.9008
Epoch 4/5
7500/7500 [=====] - 20s 3ms/sample - loss: 0.1672 - accuracy: 0.9389 - val_loss: 0.2610 - val_accuracy: 0.8964
Epoch 5/5
7500/7500 [=====] - 20s 3ms/sample - loss: 0.1487 - accuracy: 0.9475 - val_loss: 0.2686 - val_accuracy: 0.8936
Out[220]: <tensorflow.python.keras.callbacks.History at 0x192f5f62048>
```

Fig 3. Code to train the model

Algorithm	Accuracy
Random Forest	89.28%
MultinomialNB	90.84%
SVM	76.68%
KNN	81.44%
LSTM	87.56%
Bi-Directional LSTM	89.36%

Fig 4. Accuracy obtained using different models

# Project Presentation – Contribution to the Project

- I was assigned to a group with the project on Natural Language Processing, particularly Sentiment Analysis of Restaurant Reviews.
- Did the frontend work of the project and also helped a little with the backend of the project by trying with the algorithms such as SVM and KNN and seeing if the accuracy provided by these algorithms is suitable for the project.
- Did debugging.



# Outcomes - Technical

- Importance of Data and Data Preparation in Machine Learning.
- Implementing different types of ML algorithm and applications.
- Understood the mathematics behind Machine Learning.
- Learned the application of Natural Language Processing.



# Outcomes – Non Technical

- Teamwork.
- Confidence.
- Skills.
- Organizing the task.
- Open Mindedness
- Clarity.



# Conclusion

- The guidance provided by Mr. Gurvansh Singh helped me learn new technologies and complete my internship with ease. I gained valuable work experience and the internship also made me realize how to handle the responsibilities and execute the given tasks within the deadlines.
- To me it was beneficial for work in a dedicated team, to exchange knowledge, to learn new tools and new technologies.
- During this internship I learned how to perform data pre-processing, different libraries in Python programming language that support machine learning, applications of machine learning, additionally I was also taught the types of machine learning algorithms including supervised, unsupervised and reinforcement learning and their working.
- Our guide at the company was friendly and supported my team at every stage of our project and was quick to solve our doubts.



**Thank You !**