A hand-drawn wireframe of a news article sorting application is shown in a spiral notebook. The wireframe includes a header section, a sidebar with a search bar and filters, a main content area with a list of articles, and a footer. The title '#Content' is written in large, bold, black letters at the top right of the page. The text 'Detailed Project Report' and 'News Article Sorting' is overlaid in white, centered on the page. The author's name 'Rajat Chaudhari' is written in black below the title. A hand is visible at the bottom right, holding a pencil and drawing a line on the wireframe. A keyboard is visible in the background.

# #Content

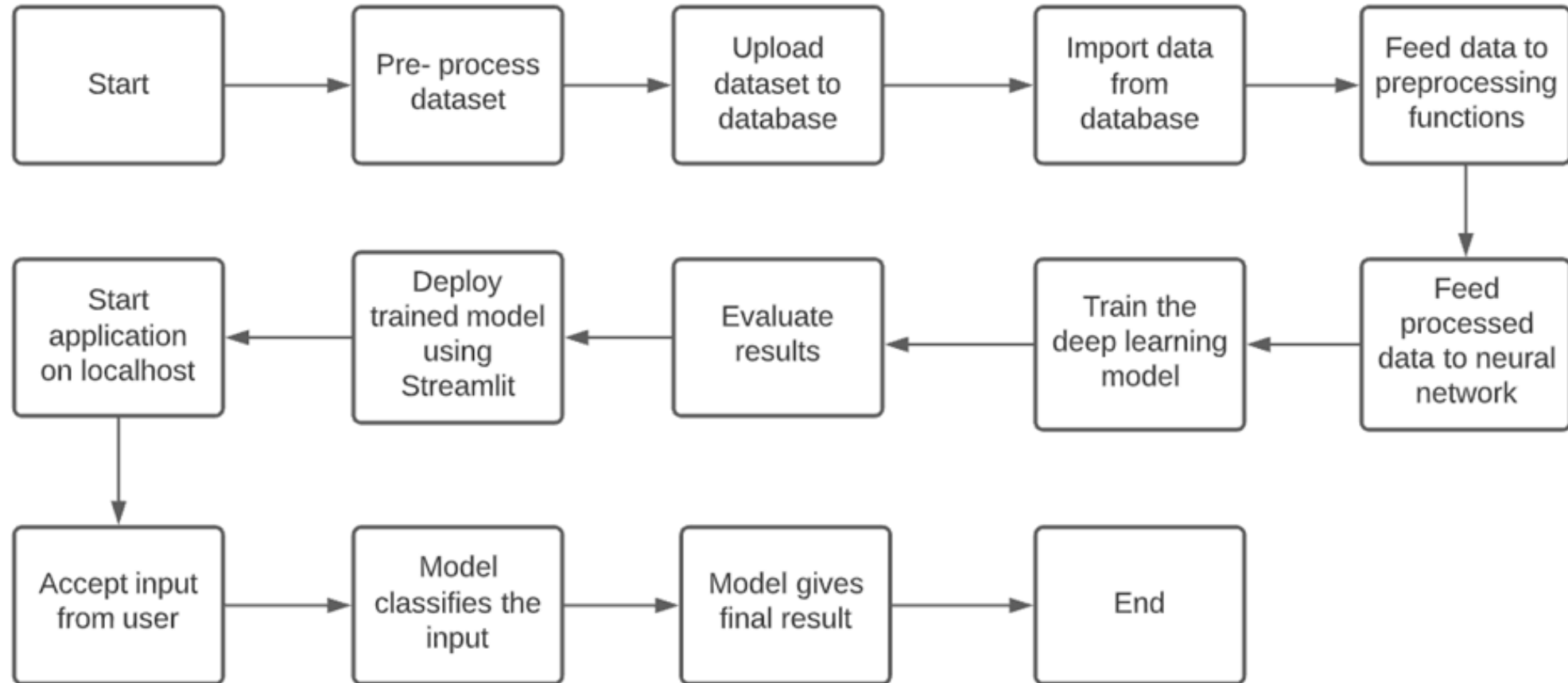
## Detailed Project Report News Article Sorting

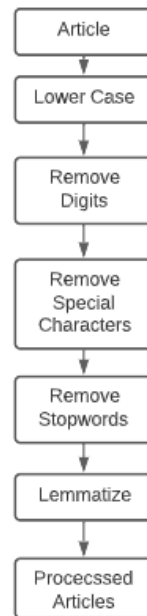
Rajat Chaudhari

# Objective

- To propose and implement a system that classifies a news article in one of the five classes (business, technology, sport, entertainment, politics).
- The system should provide a User Interface (UI) for users to use the classification model.

# Architecture





# Data Transformation

- Data transformation was performed on the data before inserting it into the database, the following operations were performed on the articles:

# Data Insertion

- The transformed data is then loaded into a CSV file.
- The CSV file is then inserted in a Cassandra Database using a python script.
- The data insertion and update of the database is automated using python scripts.

# Dataset

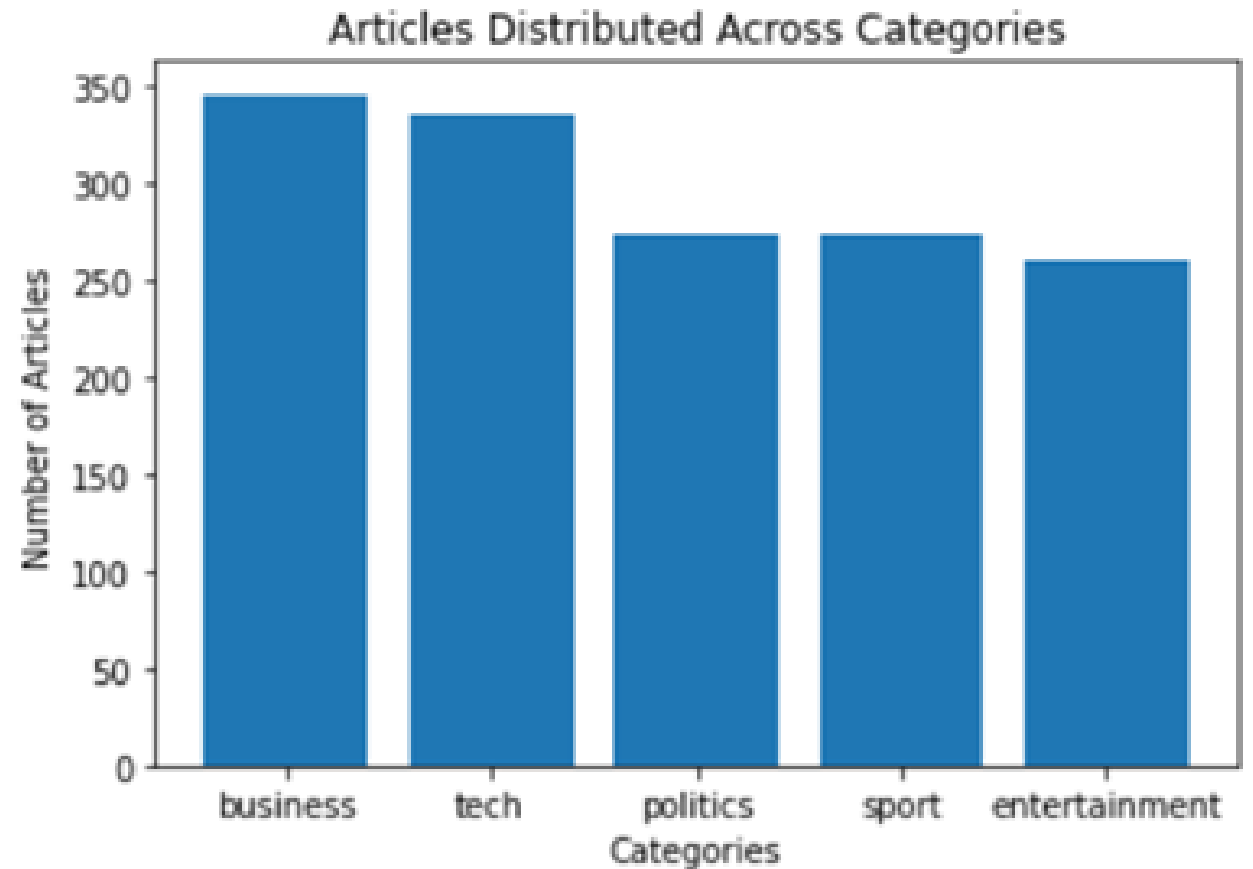
The data consists 1490 records for training set, 736 records for test set. The data fields in the dataset are:

- ArticleID
- Article
- Category

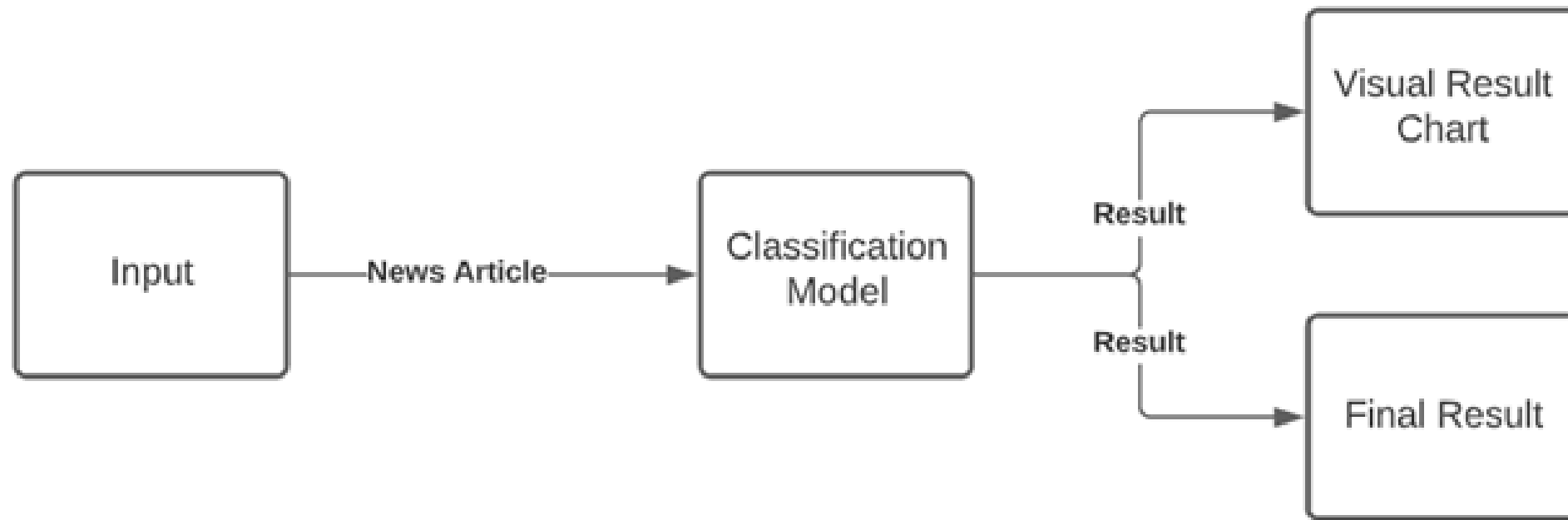
	article_id	news	category
0	1584	alicia keys to open us super bowl r&b star ali...	entertainment
1	1863	johnson uncertain about euro bid jade johnson ...	sport
2	1765	nintendo ds aims to touch gamers the mobile ga...	tech
3	2062	turkey deal to help world peace a deal bring...	politics
4	1199	sales fail to boost high street the january ...	business

# Dataset

- The articles are distributed across the 5 categories in the following way.

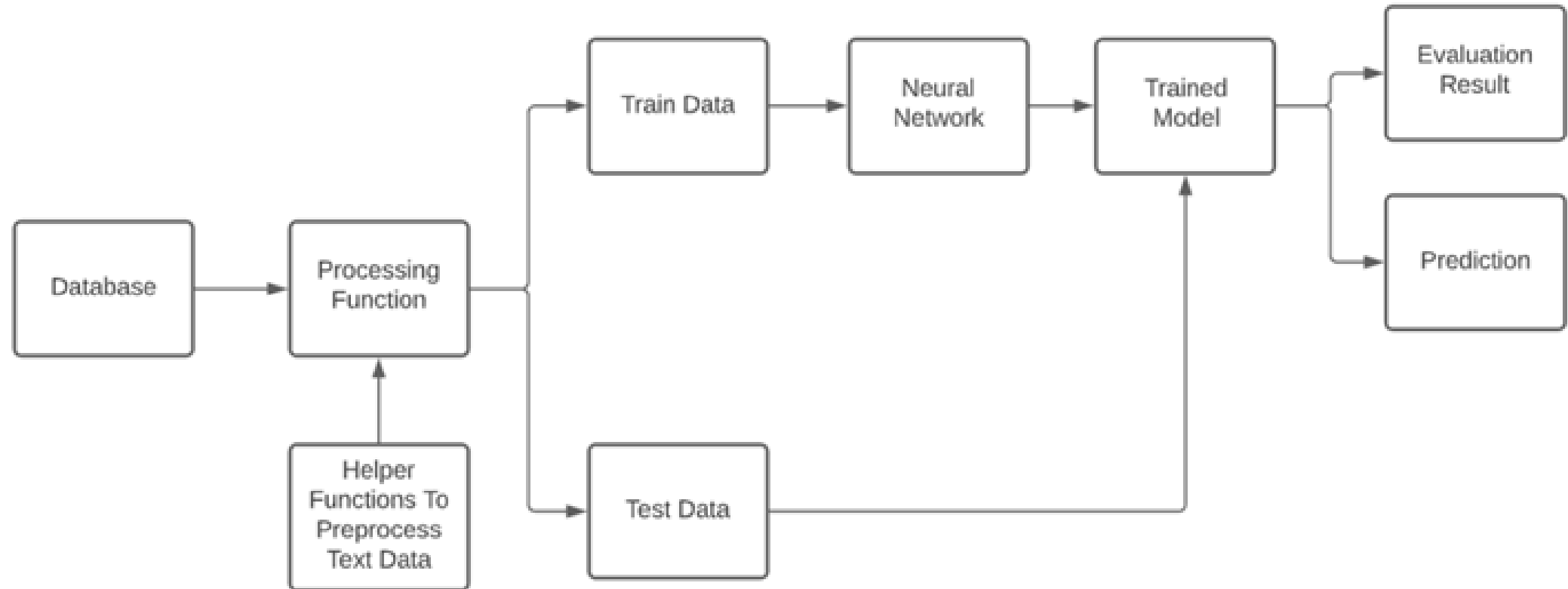


# Process Flow





# Model Training



# Model Validation

K Fold Cross Validation method was used to validate the performance of the model.

Total number of folds used were 4.

The final test accuracy of the model after cross validation is 96.03%

# Deployment

- The model was deployed using Streamlit python package.
- The model can be accessed here, <https://share.streamlit.io/rajatrc1705/news-article-sorting/main/main.py>
- The model predicts the class to which the news article belongs.



# Logging

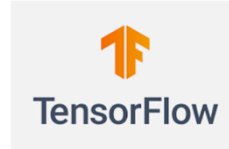
Python library logging is used to log different kinds of messages to a log file.

There are 5 levels of logging messages in the logging python library, namely:

- DEBUG
  - INFO
  - WARNING
  - ERROR
  - CRITICAL
- For this project, the level is set to INFO. All the logging messages are written to logs.txt file.

# Technology Stack

- Streamlit
- Python
- Tensorflow
- Keras
- Nltk
- Jupyter Notebook



# Conclusion

- This project provides a User Interface to a deep learning classifier model. The project is capable of accepting a news article input from the user in the form of a text. The model then processes the article and then generates a bar chart graph. The bar chart graph displays the percentage of which category the model thinks the news article belongs to. The system also displays a final result classifying the news article in one of the 5 categories (business, entertainment, politics, sport, technology).