**IST\_664\_NATURAL\_LANGUAGE\_PROCESSING.**

Dear Professor,

Please find below the **order** and **instructions** to access and review the files inside the project folder:

**1. Start with the Final Project Report**

* **File:** IST\_664\_FINAL\_PROJECT\_REPORT.docx
* **Description:**  
  This is the **main final project report** titled *"Sentiment Analysis of Movie Reviews Using Kaggle Dataset."*  
  It provides a full explanation of the dataset, data preprocessing, feature engineering, model building, evaluations, and conclusions.  
  **Please review this first** to get an overall understanding of the project.

**2. Refer to the Final Report in PDF Format (Backup Copy)**

* **File:** IST\_664\_FINAL\_PROJECT\_REPORT.pdf
* **Description:**  
  This is the **PDF version** of the same final project report.

**3. Review the Python Code Files Supporting the Project**

These Python scripts contain all the feature extraction, processing, and model-building logic used:

|  |  |  |
| --- | --- | --- |
| **Order** | **File Name** | **Purpose** |
| 3.1 | classifyKaggle.py | Main classification pipeline: preprocessing, feature extraction, and training models. |
| 3.2 | crossval.py | Script for cross-validation (precision, recall, F1-score) with Naive Bayes. |
| 3.3 | sentiment\_read\_LIWC\_pos\_neg\_words.py | LIWC Positive/Negative word list extraction utility. |
|  |  |  |
| 3.4 | sentiment\_read\_subjectivity.py | Subjectivity Lexicon Reader (for SL features extraction). |

* Start with classifyKaggle.py to understand the entire pipeline.
* Check supporting scripts (crossval.py, LIWC, SL files) to see how external features are loaded and used.

|  |  |
| --- | --- |
| **Purpose** | **Software Needed** |
| **Reading Reports** | **Microsoft Word / Google Docs / LibreOffice** |
| **Viewing PDF** | **Adobe Reader / Browser PDF Viewer** |
| **Running Python Code** | **Python 3.6+ Environment** |
| **Python Libraries** | **nltk installed** |

**Summary**

* **Step 1:** Read the final project report (.docx or .pdf).
* **Step 2:** Review the main code pipeline.
* **Step 3:** Refer to LIWC and Subjectivity lexicon scripts if needed.

This order will ensure a smooth understanding of the project flow from concept to execution.

**Thank you!**

**GITHUB:-** <https://github.com/MIHIR0501/Sentiment-analysis-of-movie-reviews-using-NLP.>