AI Assignment 3

BINARY AND MULTI-CLASS CLASSIFICATION FROM TEXT – MACHINE LEARNING CYCLE

By

Team IOIT

AI Assignment 3

AI 2024 Online Summer Internship

Make Your Place in Top 20% AI Engineers of the World

Sentiment Prediction from Text
(Machine Learning Cycle)

Quick Recap – AI Assignment 2

AI Assignment 2 focussed on RIGHT Hand Side (RHS), i.e., Model + Error, of the following Equation, when FEATURES are ALREADY EXTRACTED.:
Data = Model + Error

Aim

The focus of this Assignment is on Right Hand Side (RHS) of the Equation (when FEATURES are EXTRACTED from Text) i.e.,
Model + Error

The aims of this Assignment are:

- **To introduce how to develop Machine Learning based Real-world Applications using NLTK, Python, Pandas and Scikit-learn**
- **To introduce how we can develop Machine Learning Models by using the Data in which FEATURES are EXTRACTED from Text (for both Binary and Multi-class Classification Tasks)**
- **To introduce how we can execute the Machine Learning Cycle (Training Phase, Testing Phase, Application Phase and Feedback Phase) to develop Real-world Applications**

Learning Outcomes

After reading, understanding, analyzing and doing this Assignment, In Sha Allah, you will learn:

- **How to use NLTK, Python, Pandas and Scikit-learn to build Machine Learning based Real-world Application?**
- **How to execute Machine Learning Cycle (Training Phase, Testing Phase, Application Phase and Feedback Phase) to develop Real-world Applications using the Data in which FEATURES are EXTRACTED from Text (for both Binary and Multi-class Classification Tasks)?**

Tasks

Note that this Assignment is planned, designed, and developed using a HalfCooked Approach i.e., some work is done by the instructor and the remaining will be done by you. Since the Assignment is to develop Machine Learning based Realworld Applications using NLTK, Python, Pandas and Scikit-learn. Following the HalfCooked Approach, Instructor has provided with the Code, Data, and Documentation, for the Project titled: Sentiment Prediction from Text. Note that the Titanic Project is developed using Train-Test Split Approach. Your job is to carry out the Tasks given below and answer the associated questions.

Your main job is to Code, Document and Develop Prototype System for the TODO Task, i.e., Gender Identification from Text, given in the documentation of Titanic Passenger Survival Prediction Project by performing the following Tasks.

- **Task 1**
  - **Run the Code for Sentiment Prediction from Text (using Train-Test Split Approach)**
- **Task 2**
  - **Edit the Code given for Sentiemnt Prediction from Text, so that it works for Gender Identification from Text (Binary Classification from Text)**
  - **Note**
    - **Only select 100 instances (50 Male, 50 Female) from gender_identificaiton_dataset File**

- **Task 3**
  - **Edit the Code given for Sentiemnt Prediction from Text, so that it works for Emotion Prediction from Text (Multi-class Classification)**
    - **See emotion_prediction_dataset File**

Guidelines for Assignment Submission

Deadline: 28-Jul-2024 (23:59 PST)

How to Submit?
Upload your Tutorial (Assignment Solution) on Microsoft Teams. Upload the following files:
1. Jupyter Notebook File
2. Jupyter Notebook in HTML Format
3. Dataset(s)