

Project 10

Intensity Analysis (Build your own model using NLP and Python)

The objective of this project is to develop an intelligent system using NLP to predict the intensity in the text reviews. By analyzing various parameters and process data, the system will predict the intensity where its happiness, angeriness or sadness. This predictive capability will enable to proactively optimize their processes, and improve overall customer satisfaction.

Focus Areas:

- 📅 Data Collection: Gather all the intensity data, including the text and its intensity.
- 📅 Data Preprocessing: Clean, preprocess, and transform the data to make it suitable for machine learning models.
- 📅 Feature Engineering: Extract relevant features and identify key process variables that impact intensity.
- 📅 Model Selection: Choose appropriate machine learning algorithms for intensity classification.
- 📅 Model Training: Train the selected models using the preprocessed data.
- 📅 Model Evaluation: Assess the performance of the trained models using appropriate evaluation metrics.
- 📅 Hyperparameter Tuning: Optimize model hyperparameters to improve predictive accuracy.
- 📅 Deployment: Deploy the trained model in a production environment for real-time predictions.

Your focus in this project should be on the following:

The following is recommendation of the steps that should be employed towards attempting to solve this problem statement:

- ✦ **Data Collection:** Collect and preprocess data for natural language processing (NLP) tasks. This may include text data for training the language model.
- ✦ **Model Development:** Use machine learning techniques such as deep learning to develop models for NLP. These models should be trained on the collected data and fine-tuned for improved accuracy.
- ✦ **Testing and Validation:** Test the classification in different scenarios and environments to ensure it can analyze appropriately. The performance of the model should be evaluated, and any issues should be addressed.

Timeline

We expect you to do your best and submit a solution within 2 weeks.

Deliverables

Please share the following deliverables in a zip file.

- 📁 A report (PDF) detailing:
- 📁 Description of design choices and Performance evaluation of the model
- 📁 Discussion of future work
- 📁 The source code used to create the pipeline

Tasks/Activities List

Your code should contain the following activities/Analysis:

- 📁 Collect the data from any possible resources.
- 📁 Data Preprocessing.
- 📁 Feature Engineering and feature selection.
- 📁 Train/Test Split
- 📁 Choose the metrics for the model evaluation
- 📁 Model Selection, Training, Predicting and Assessment
- 📁 Hyperparameter Tuning/Model Improvement
- 📁 Model deployment plan.

Success Metrics

Below are the metrics for the successful submission of this case study.

- 📁 The accuracy of the model on the test data set should be > 85% (Subjective in nature)
- 📁 Add methods for Hyperparameter tuning.
- 📁 Perform model validation.

Bonus Points

- 📁 You can package your solution in a zip file included with a README that explains the installation and execution of the end-to-end pipeline.
- 📁 You can demonstrate your documentation skills by describing how it benefits our company.

Data:

The dataset for this project can be accessed by clicking the link provided below

[Intensity_data.zip](#)