INTEL – UNNATI INDUSTRIAL TRAINING PROGRAM
2024-2025

PROJECT NAME

PS-11 Intel Products Sentiment Analysis from Online Reviews

TEAM NAME : UNITY CIRCLE

TEAM MEMBER 1 : KIRTHANA MOHAN R

TEAM MEMBER 2 : SWETHA S

TEAM MEMBER 3:_KRITHIKA S

COLLEGE NAME : SRI SAI RAM INSTITUTE OF TECHNOLOGY

PROBLEM STATEMENT

- Customer feedback and Reviews plays a crucial role in the continuous improvement of products and services. Companies often receive vast amounts of textual reviews from end users and tech reviewers on various platforms, making it challenging to manually analyse and derive actionable insights from this data.

- The problem at hand is to analyze and interpret user sentiments and feedback related to Intel's 12th, 13th, and 14th generation processors, in order to understand user experiences, identify common issues, and derive actionable insights for improving future product iterations.

UNIQUE IDEA BRIEF

- To solve the problem of analyzing user sentiments and feedback on Intel's processors, this project uses several machine learning models, including Random Forest, XGBoost, Logistic Regression, SVM, CNN, and LSTM, to classify and predict sentiments from user reviews. The project also uses TF-IDF vectorization to extract features from text data. RandomizedSearchCV is applied to find the best model parameters.
- Additionally, a chatbot is developed using the Gemini API, which provides users with detailed information about Intel processors, user experiences, pricing trends, and sentiment analysis. This chatbot can answer questions, give recommendations, and help users make informed decisions based on the analyzed data.

FEATURES OFFERED

•**Sentiment Analysis**: The project performs sentiment analysis on user reviews using various models like Random Forest, XGBoost, Logistic Regression, SVM, CNN, and LSTM, with Random Forest showing the best performance.

•**Detailed Descriptions**: It provides comprehensive descriptions of Intel's 12th, 13th, and 14th generation processors, including their specifications and key features.

•**User Experience Insights**: By analyzing user reviews, the project offers valuable insights into common user experiences and satisfaction levels.

•**Recommendations**: Based on sentiment analysis and data trends, the project suggests improvements to enhance processor performance and user satisfaction.

PROCESS FLOW

1 Data Collection:
•Gather user reviews and comments about Intel's 12th, 13th, and 14th generation processors from various sources.

2 Data Preprocessing:
•Clean the data by handling missing values and removing non-text elements.
•Translate foreign language reviews to English if necessary.
•Convert the date format to a consistent format using pd.to_datetime

•**Exploratory Data Analysis (EDA)**:
•Generate visualizations such as SKU distribution, word clouds, and correlation matrices to understand data patterns.
•Identify common words and complaints from the reviews.

4 Feature Extraction:
•Use TF-IDF vectorization to convert text data into numerical features suitable for machine learning models.

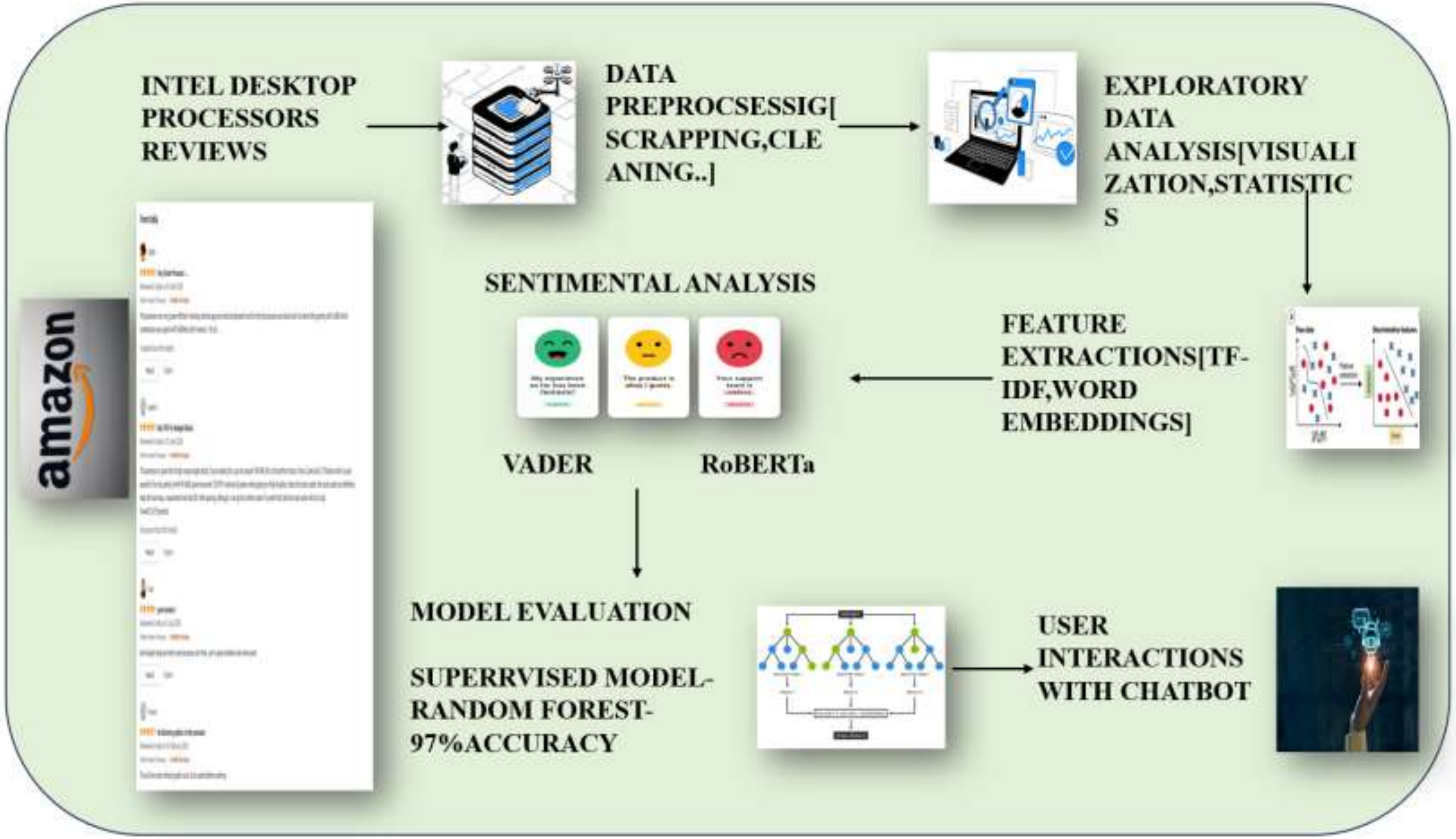
5 Sentiment Analysis:
•Apply sentiment analysis techniques using models like VADER and RoBERTa to classify the sentiments of the reviews.

6 Model Training and Evaluation:
•Split the data into training and testing sets.
•Train multiple machine learning models (Random Forest, XGBoost, Logistic Regression, SVM, CNN, LSTM) on the training data.
•Use RandomizedSearchCV for hyperparameter tuning of the Random Forest model.
•Evaluate the performance of each model using metrics like accuracy on the test data.

8 Chatbot Integration:
•Develop a chatbot using the Gemini API.
•Enable the chatbot to provide detailed information on Intel processors, user experiences, pricing trends, sentiment analysis, and recommendations.

9 Results and Conclusions:
•Summarize the performance of different models.
•Highlight key findings and insights from the sentiment analysis.
•Provide recommendations for future products based on user reviews and sentiments.

ARCHITECTURE DIAGRAM



TECHNOLOGIES USED

1.Rule-based methods

- **VADER (valence aware dictionary and sentiment reasoner):**
- Technique**:** Rule-based model that uses a lexicon of sentiment-related words.

2. Machine Learning approaches

- **Support Vector Machines (SVM):**
- Technique**:** Supervised learning model that finds the hyperplane that best separates different classes in feature space.
- **Random Forest:**
- Technique**:** Ensemble learning method using multiple decision trees.
- **Logistic Regression**:
- Technique**:** Statistical model for binary classification

- **XGBoost:**

- Technique: Ensemble learning method using gradient boosting.

3. Deep Learning models

- **Convolutional Neural Networks (CNN):**

- Technique**:** Neural networks traditionally used for image processing, adapted for text by treating sentences as sequences of word vectors.

4. unsupervised learning approaches

- **K mean clustering**

TOOLS USED :

1 NLTK: For text preprocessing and tokenization.

2 Scikit-learn: For machine learning model training and evaluation, including:

- **RandomForestClassifier**: A machine learning algorithm for classification.
- **LogisticRegression**: Another classification algorithm.
- **SVM**: Support Vector Machine for classification tasks.
- **TfidfVectorizer**: For converting text data into numerical features.
- **RandomizedSearchCV**: For hyperparameter tuning.

3 Gensim: For word embedding and Word2Vec model.

4 TensorFlow/Keras: For deep learning models, including:

- **CNN**: Convolutional Neural Network.
- **LSTM**: Long Short-Term Memory network.

5 VADER: A lexicon and rule-based sentiment analysis tool.

6 Hugging Face Transformers: For advanced sentiment analysis models like RoBERTa.

7 Matplotlib and **Seaborn**: For generating visualizations and heatmaps.

8 WordCloud: For generating word cloud visualizations.

9 Gemini API: For integrating the chatbot functionality to respond to user inquiries about Intel processors.

TEAM MEMBERS AND CONTRIBUTION:

- The team collaborated on various aspects of the project, with members specializing in data collection, preprocessing, model development, and evaluation. Their combined efforts ensured comprehensive coverage of the workflow, resulting in a robust sentiment analysis system.

- 1.KIRTHANA MOHAN R -

- web scrapping ,
- created LSTM , ROBERTA , VADER model for 12th gen processor and 13th gen processor
- drafted the final report

- 2. SWETHA S –

- web scrapping
- created CNN , RANDOM FOREST , XGBOOST , LOGISTIC REGRESSION and SVM model for 12th gen processor and 13th gen processor
- Created the CHATBOT for responding to inquiries regarding Intel's 12th, 13th, and 14th generation processors

- 3. KRITHIKA S

- Web scrapping
- created CNN , RANDOM FOREST , XGBOOST , LOGISTIC REGRESSION and SVM model for 14th gen processor

CONCLUSION

- The evaluation of different machine learning models on the 12th, 13th, and 14th generation Intel processors reveals that the Random Forest classifier consistently outperforms other models, achieving the highest accuracy rates of 82% for the 12th and 14th generations and an impressive 97% for the 13th generation. Other models such as XGBoost, Logistic Regression, SVM, CNN, and LSTM show varied performance, with XGBoost and Logistic Regression generally performing better than SVM, CNN, and LSTM. The significant drop in accuracy for most models on the 13th generation, except for Random Forest, indicates potential variability in the dataset or model-specific performance issues. The chatbot leveraging this data provides users with comprehensive insights, ranging from detailed processor descriptions to sentiment analysis based on user reviews. These findings suggest that the Random Forest model is the most reliable for predicting user sentiment and performance trends across different Intel processor generations.

- This sentiment analysis project involved scraping Amazon product reviews, preprocessing the text data, and training an LSTM model to classify sentiments as Positive, Neutral, or Negative. The main findings indicate that the model accurately identifies sentiment trends, revealing valuable insights into customer opinions and product performance. These implications suggest potential enhancements in customer feedback analysis, helping businesses improve products and services based on user sentiment.