**Fake Review Detection and Topic Modelling**

**1.Introduction:**

In the age of digital commerce, reviews play a crucial role in shaping customer trust and purchase decisions. However, the proliferation of **fake reviews** has undermined this trust, making it essential to identify and filter out deceptive content. Simultaneously, extracting meaningful **topics** from reviews can provide actionable insights for businesses and consumers.

* 1. **PROJECT OBJECTIVES:**

This project is divided into two major components:

1. **Fake Review Classification:**
   * This involves using machine learning and deep learning models to identify whether a review is genuine or fake.
   * Models like **LSTM**, **CNN** **and CNN-LSTM** were explored to achieve high accuracy in classification.
2. **Topic Modelling:**
   * This part focuses on understanding the core themes and topics discussed in the reviews.
   * Techniques such as **Latent Dirichlet Allocation (LDA)**, **Non-Negative Matrix Factorization (NMF)**, and clustering algorithms like **K-Means** and **DBSCAN** were used.
   * Sentiment analysis using **BERT** helped refine the understanding of customer feedback.
   1. **Dataset Description**

The dataset contains reviews labeled as fake or original, along with the rating and product category for each review. The target column is labeled as "label."

**2. Part I: Fake Review Classification**

**2.1 Data Preprocessing:**

* Steps taken: Data cleaning, handling missing values, text preprocessing.
* Tokenization and padding for deep learning models.

**2.2 Models Used:**

**MACHINE LEARNING MODELS**

|  |  |  |
| --- | --- | --- |
| **S.NO** | **MODELS** | **ACCURACY** |
| 1 | Logistic Regression | 86.32% |
| 2 | Multinomial Naïve Bayes | 84.43% |
| 3 | Support Vector Machine | 87.14% |
| 4 | Random Forest Classifier | 84.39% |
| 5 | Decision Tree | 73.72% |
| 6 | K-Nearest Neighbour | 65.52% |

**DEEP LEARNING MODELS**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.NO** | **MODELS** | **TEST LOSS** | **TEST ACCURACY** |
| 1 | LSTM | 0.297 | 0.8958 |
| 2 | CNN | 0.6369 | 0.8872 |
| 3 | CNN-LSTM | 0.3538 | 0.8976 |

**Best Model Justification**

1. Machine Learning Models:
   * **Support Vector Machine (SVM)** emerged as the top performer among traditional models with an accuracy of **87.14%**.
   * SVM's ability to handle high-dimensional data and its robustness in separating non-linearly separable classes made it the most suitable choice among ML models.
2. **Deep Learning Models:**
   * The **CNN-LSTM** model achieved the highest test accuracy of **89.76%** and demonstrated a balance between test loss (**0.3538**) and accuracy.
   * CNN's capability to capture spatial patterns combined with LSTM's strength in learning sequential dependencies made this hybrid model ideal for analyzing text data.

**Preferred Model: CNN-LSTM**

* While both **LSTM** and **CNN-LSTM** demonstrated high accuracy, the slight edge in performance for CNN-LSTM indicates its superior ability to generalize.
* **Justification:**
  + CNN extracts critical features from text at different granular levels.
  + LSTM handles sequential data effectively, capturing context and dependencies.
  + The combined architecture provides an optimal balance between feature extraction and temporal understanding.

For fake review classification, **CNN-LSTM** is the most effective model, leveraging the strengths of both convolutional and recurrent networks for accurate and reliable predictions.

**3. Part II: Topic Modelling And Insights**

**3.1 Data Preprocessing**

* Tokenization, stopword removal, lemmatization.

**3.2 Techniques Used**

|  |  |  |
| --- | --- | --- |
| **S.No** | **TECHNIQUE** | **COHERENCE** |
| 1 | LDA | 0.5813 |
| 2 | NMF | 0.6852 |

**Observation**:

* NMF (Non-negative Matrix Factorization) outperforms LDA (Latent Dirichlet Allocation) in terms of coherence score, indicating better topic interpretability and alignment with human understanding.
* Higher coherence in NMF suggests that the generated topics are more semantically meaningful and consistent.

|  |  |  |
| --- | --- | --- |
| **S.No** | **TECHNIQUE** | **SILHOUETTE SCORE** |
| 1 | K-MEANS | 0.4461 |
| 2 | DBSCAN | -0.0381 |

* **Observation**:
  + K-Means achieves a positive silhouette score (**0.4461**), signifying moderate clustering performance.
  + DBSCAN yields a negative silhouette score, suggesting that the clustering boundaries are not well-defined for this dataset.

**BERT**

* Sentiment analysis of extracted topics.

**4. Challenges and Improvements**

1. **Runtime Limitations**:
   * Implementing BERT for sentiment analysis presented significant runtime challenges, especially in the Google Colab environment, where the runtime would occasionally terminate before processing was complete due to the extensive computational requirements.
   * **Improvement**: To address this, we explored optimizing data processing by batching inputs and leveraging hardware accelerators (TPUs or GPUs) to reduce execution time.
2. **Model Comparisons and Accuracy Trade-offs**:
   * While **traditional machine learning models** like Support Vector Machine (SVM) showed strong performance (87.14%), deep learning models such as **CNN-LSTM** outperformed them with higher accuracy (89.76%) and lower test loss (0.3538).
   * **Improvement**: Deep learning models were fine-tuned with hyperparameter optimization and additional training epochs to maximize performance.
3. **Topic Modeling Coherence:**
   * **Challenge**: Latent Dirichlet Allocation (LDA) generated topics with a moderate coherence score (0.5813), which indicated some limitations in topic interpretability.
   * **Improvement**: By adopting Non-Negative Matrix Factorization (NMF), we achieved a higher coherence score (0.6852), demonstrating more meaningful and interpretable topics.
4. **Clustering Performance**:
   * **Challenge**: Clustering techniques like DBSCAN produced poorly defined clusters with a negative silhouette score (-0.0381), highlighting its unsuitability for this dataset.
   * **Improvement**: K-Means clustering, with a silhouette score of 0.4461, provided better-defined clusters. Future improvements may involve exploring hierarchical clustering or fine-tuning DBSCAN parameters.
5. **Balancing Computational Efficiency with Model Performance**:
   * **Challenge**: Some models (e.g., CNN-LSTM) required higher computational resources, while simpler models (e.g., Logistic Regression) offered faster execution at a slight trade-off in accuracy.
   * **Improvement**: We prioritized models like CNN-LSTM for critical tasks while retaining simpler models for scenarios demanding quicker predictions.

**PRACTICAL USE CASES OF THE PROJECT:**

**FAKE REVIEW CLASSIFICATION**

* **E-commerce Platforms**
* Digital Marketing
* **Consumer Protection Agencies**

**TOPIC MODELLING**

* **Customer Insights and Sentiment Analysis**
* **Content Moderation**
* **Personalized Recommendations**
* Fraud Detection Systems
* Industry Reports
* **Behavioral Studies**

**Future Directions:**

To further enhance the effectiveness of the models and their applicability, the following steps can be considered:

* Incorporating **ensemble techniques** to combine the strengths of multiple models.
* Expanding the dataset to include diverse domains and languages.
* Optimizing computational efficiency to handle larger datasets and complex models like BERT.

**CONCLUSION:**

This project underscores the transformative potential of AI in extracting actionable insights, fostering transparency, and building trust in digital platforms, paving the way for future innovations in text analytics.