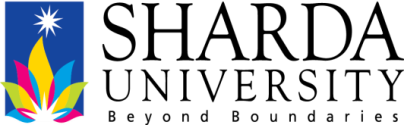
a

****

**PROJECT BASED LEARNING (PBL-2) LAB (CSP297)**

**Enhancing Review Authenticity:**

**Deploying Machine Learning Techniques for Fake Review Detection**

**B.TECH 2nd YEAR**

**SEMESTER: 4th**

**SESSION: 2023-2024**

**Submitted By:**

**Mehak Parashar (Student ID)**

**Shivank Mehrotra (Student ID)**

**Suraj Belwal (2022434098)**

**SECTION: J**

**Submitted To**

**Dr. Saptdeepa Kalita**

**Associate Professor**

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**SHARDA SCHOOL OF ENGINEERING & TECHNOLOGY**

**SHARDA UNIVERSITY, GREATER NOIDA**

Table of Contents Page No

[Project Title 3](#_Toc163992389)

[Team / Group Formation: 3](#_Toc163992390)

[Technologies to be used 3](#_Toc163992391)

[Tools 4](#_Toc163992392)

[Problem Statement 4](#_Toc163992393)

[Literature Survey 5](#_Toc163992394)

[Project Description 6](#_Toc163992395)

[Project Modules: Design/Algorithm 6](#_Toc163992396)

[Implementation Methodology 6](#_Toc163992397)

[Result & Conclusion 7](#_Toc163992398)

[Future Scope and Further Enhancement of Project 8](#_Toc163992399)

[Advantages of this Project 8](#_Toc163992400)

[Outcome 9](#_Toc163992401)

[References 9](#_Toc163992402)

# Project Title

Enhancing Review Authenticity: Deploying Machine Learning Techniques for Fake Review Detection

# Team / Group Formation:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S. No | Student Name | Roll Number | System ID | Role |
| 1 | Mehak Parashar |  |  |  |
| 2 | Shivank Mehrotra |  |  |  |
| 3 | Suraj Belwal | 2201010727 | 2022434098 |  |

# Technologies to be used

Software Platform:

a) Front-end:

HTML/CSS/JavaScript for web interface development

IDEs like Visual Studio Code or Sublime Text for development

Hardware Platform:

RAM: Sufficient RAM to handle the computational requirements of the machine learning models and web application.

Hard Disk: Adequate storage space for storing datasets, model files, and project resources.

OS: Compatible operating system such as Windows, macOS, or Linux for development and deployment.

Editor: Text editors or Integrated Development Environments (IDEs) for writing and debugging code.

Browser: Web browsers like Google Chrome, Mozilla Firefox, or Safari for testing and running the web-based application.

# Tools

1. Python: Python served as the primary language for writing the project code.

Version: 3.8

Vendor: Python Software Foundation

1. Pandas: A Python library used for data manipulation and analysis, particularly useful for handling structured data.

Version: 1.2.4

Vendor: Pandas Development Team

1. Matplotlib: A plotting library for Python and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications.

Version: 3.3.4

Vendor: John D. Hunter, Matplotlib Development Team

1. NLTK (Natural Language Toolkit): A leading platform for building Python programs to work with human language data (text). It was used for text preprocessing, including tokenization and removing stop words.

Version: 3.5

Vendor: NLTK Project

1. Scikit-learn: A machine learning library for Python. It features various classification, regression, and clustering algorithms, and was used for model training, evaluation, and validation.

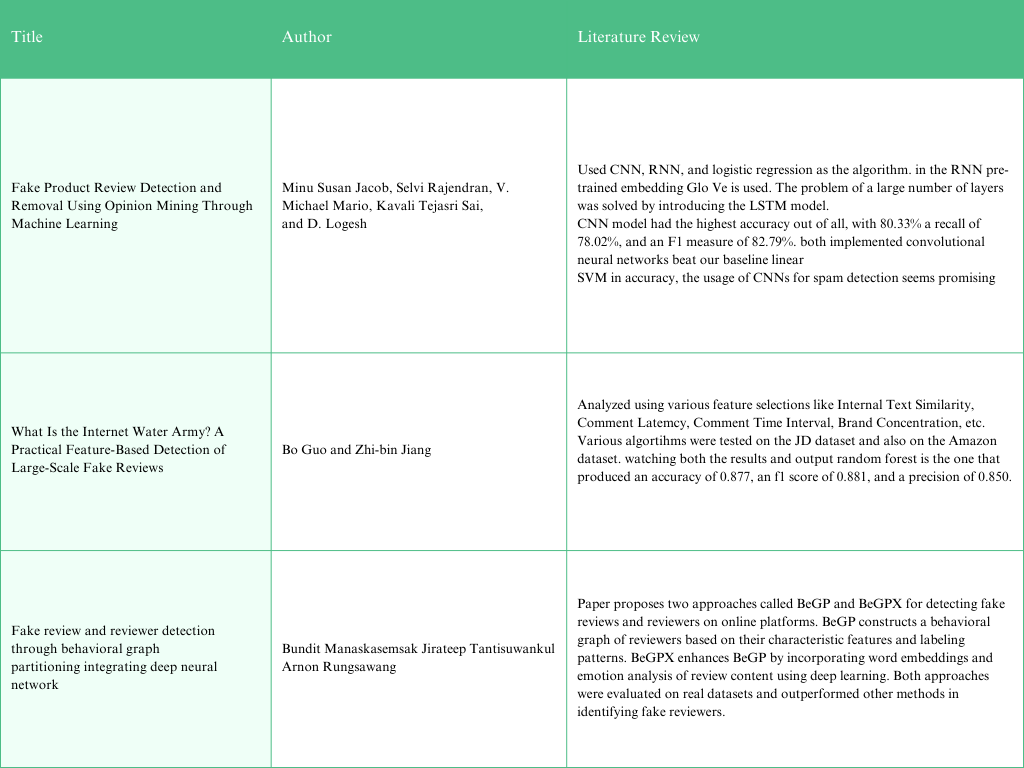
Version: 0.24.1

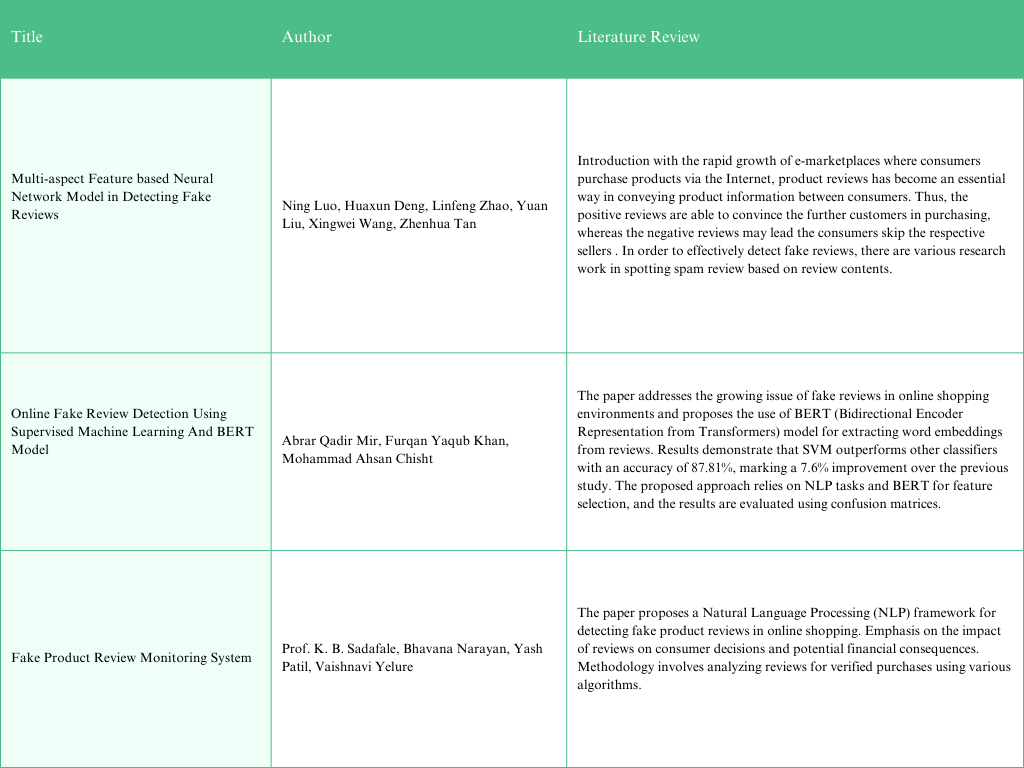
Vendor: Scikit-learn Developers

# Problem Statement

In today’s digital world, the role that genuine reviews play towards the audience is massive, and with some companies employing deceptive or misleading reviews on their products to manipulate public opinion is a matter of concern. This research addresses the unavoidable task of fake product review detection by using advanced natural language processing techniques. The implications of this research extend to multiple diverse online platforms, particularly e-commerce websites, where the proposed approach can be integrated to automatically identify as well as flag suspicious reviews. By enhancing the trustworthiness of user-generated content, this system empowers customers with more reliable and honest information, developing a transparent and genuine online shopping experience.

# Literature Survey



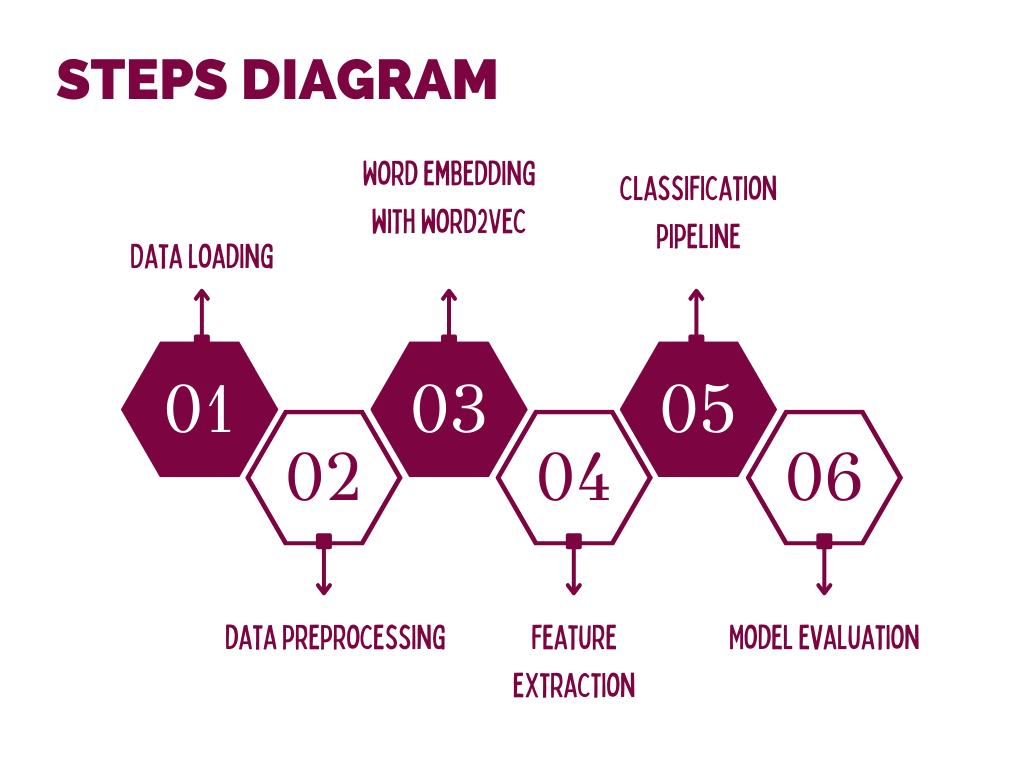


# 

# Project Description

The research deeply examines the detailed approach of pre-processing steps, highlighting the importance of tokenization, stemming, and Word2Vec feature extraction to collect and accumulate meaningful representations of text data. After running a detailed experiment and also after comparing the results of both methods, we establish that Word2Vec-based features yield higher accuracy compared to the TF-IDF and BOW methods. Furthermore, the study evaluates the effectiveness of several classification algorithms, such as Random Forest, Support Vector Machines, K-Nearest Neighbours, Decision Tree, Naive Bayes, and Logistic Regression. The results prove that our Word2Vec-enhanced models every time outperform their counterparts, exhibiting the potential for robust fake review detection.

# Project Modules: Design/Algorithm



# Implementation Methodology

1. System Design:

The fake review detection system was designed to be implemented in two primary stages: model development and browser extension integration. Given the current status of the project, the focus remains on model development, which includes data preprocessing, feature extraction, model training, and evaluation.

2. Data Preprocessing:

The raw data from a dataset of product reviews is first processed to remove noise and standardize the text. This includes removing punctuation, converting text to lowercase, tokenizing text into words, removing stop words, and applying stemming and lemmatization techniques. This step ensures that the data fed into the model is clean and normalized, improving the accuracy of the model.

3. Feature Extraction:

Using the Word2Vec model, the preprocessed text is converted into numerical vectors that represent each word. The average of these word vectors is then calculated for each review to create a fixed-length input vector for each document. This vectorization process allows the machine learning algorithms to effectively analyze and learn from the text data.

4. Model Training:

Several machine learning classifiers are trained using the extracted features. The classifiers include Random Forest, SVM, K-Nearest Neighbors, Decision Trees, Naive Bayes, and Logistic Regression. The training process involves adjusting parameters and using cross-validation to optimize performance.

5. Model Evaluation:

Each trained model is evaluated using a separate test set to ensure that the model accurately predicts fake reviews. Performance metrics such as accuracy, precision, recall, and F1-score are computed. Confusion matrices are also used to visualize the performance of each classifier.

# Result & Conclusion

1. High Accuracy: The model has demonstrated high accuracy in distinguishing between genuine and fake reviews, making it a valuable tool for businesses and consumers alike who rely on reviews for purchasing decisions.
2. Robust Preprocessing and Feature Extraction: Utilizing a combination of NLTK and Gensim's Word2Vec for text preprocessing and feature extraction provided a robust foundation for training the classifiers. This approach ensured that the nuances of natural language were effectively captured.
3. Versatile Classifier Implementation: The system was tested with multiple classifiers, including Random Forest, SVM, and Logistic Regression, each providing insights into different aspects of learning patterns from the data, thereby facilitating a comprehensive understanding of model strengths and weaknesses.
4. Conclusion: The fake review detection system encapsulates a significant stride toward ensuring the authenticity of online content. By leveraging cutting-edge machine learning techniques and natural language processing, the system provides a scalable and effective solution to combat the prevalent issue of fake reviews. While further enhancements are necessary for integration into real-time applications, the foundation laid by this project paves the way for future innovations in the domain of digital trust and security.

# Future Scope and further enhancement of the Project

1. Extension Development: Complete the development of the web browser extension to integrate the fake review detection model directly into e-commerce websites. This will allow users to check the authenticity of reviews while browsing products online.
2. Real-Time Review Analysis: Implement real-time review analysis capabilities to provide instant feedback to users about the credibility of reviews as they browse through products. This will enhance user experience and help them make more informed purchasing decisions.
3. Enhanced Model Accuracy: Continuously improve the fake review detection model by collecting more data and incorporating advanced natural language processing techniques. This will further enhance the accuracy and reliability of the system in identifying fake reviews.

# Advantages of this Project

1. Time-saving: Users no longer need to manually sift through reviews to determine their authenticity. The automated detection system streamlines the process, saving users time and effort.
2. Increased Confidence: With the assurance of reliable reviews, users can have increased confidence in their purchasing decisions, leading to higher satisfaction with their shopping experiences.
3. Fair Competition: E-commerce platforms benefit from a more level playing field as fake reviews are identified and removed. This fosters fair competition among sellers and promotes a healthier marketplace.
4. Consumer Empowerment: Empowers consumers to take control of their online shopping experiences by providing them with the tools to distinguish between genuine and fake reviews.
5. Improved Product Quality: Sellers are incentivized to maintain high product quality and service standards as fake reviews are less likely to artificially inflate their ratings. This ultimately benefits consumers by ensuring they receive quality products and services.

# Outcome

The project on detecting fake reviews can be transformed into a commercial product aimed at e-commerce platforms and online review aggregators. The developed model can be integrated as a feature or plugin within these platforms, providing them with an additional layer of review authenticity verification. This product can be marketed to businesses as a tool to enhance consumer trust, improve user experience, and maintain the integrity of their review systems. Through productization, the project can generate revenue streams through licensing, subscription, or usage-based models, thereby translating the research and development efforts into a viable commercial offering

# References

Signature

|  |  |  |  |
| --- | --- | --- | --- |
| Student Name | Student Sign | Faculty Name | Faculty Sign |
| Mehak Parashar |  |  |  |
| Shivank Mehrotra |  |
| Suraj Belwal |  |