ABSTRACT - SEMINAR

SUBMITTED BY:

MANJARI JAYAN

ROLL NO:35

INMCA S9

**FAKE REVIEW DETECTION**

The reliability of user-generated content, such as product or service reviews, has become a key concern in the age of digital information and online reviews. The detection of false reviews has attracted substantial interest due to its potential to mislead consumers and harm businesses' reputations. This session investigates the use of machine learning techniques to address the problem of false review detection. To construct a comprehensive detection system, we use a wide variety of Python packages and tools.

The key components of our approach include data collection from a SQLite database, data preprocessing involving text cleaning and feature extraction using TF-IDF (Term Frequency-Inverse Document Frequency), and the utilization of powerful machine learning algorithm, Random Forest. Our analysis extends to the creation of meaningful features such as review length, review deviation, and maximum content similarity to enhance model performance.

To address the issue of class imbalance in the dataset, we implement under-sampling techniques to ensure a balanced representation of both genuine and fake reviews. The core innovation lies in our application of semi-supervised learning, wherein we iteratively train the model on labeled data and incorporate high-confidence predictions as additional labeled data. This iterative approach is a novel way to improve the detection of fraudulent reviews.

The seminar presents a comprehensive evaluation of model performance using various metrics, including accuracy, precision, recall, and F1-score. Additionally, we provide insights into the interpretation of a confusion matrix generated by the Random Forest classifier. The seminar underscores the importance of fake review detection in maintaining the credibility of online review platforms and aiding consumers in making informed decisions.

Furthermore, this seminar showcases the practical implementation of machine learning in real-world applications and demonstrates how a combination of data preprocessing, feature engineering, and semi-supervised learning techniques can be employed to effectively address the challenge of fake review detection.

Top of Form