WEB INTELLIGENCE EXPERIMENT

EXPERIMENT: 06

NAME: Shashwat Shah

SAPID: 60004220126

BRANCH: COMPUTER ENGINEERING (C2-2)

AIM: Implement Opinion Spam Detection

THEORY:

With the surge in e-commerce, online reviews significantly influence consumer decision-making. However, not all reviews are authentic. Opinion spam refers to deceptive or fake reviews intended to mislead consumers. These may be overly positive reviews promoting a product (promotion spam) or negative ones aimed at discrediting competitors (demotion spam). Detecting and filtering out such spam has become crucial for maintaining the integrity and reliability of review systems.

Types of Opinion Spam

1. Non-review Spam: Content that doesn't review the product, such as advertisements or irrelevant text.
2. Brand-only Review Spam: Discusses the brand or seller without addressing the specific product.
3. Untruthful Review Spam: Fabricated or exaggerated reviews, either extremely positive or negative, written with the intent to mislead.
4. Duplicate Reviews: Identical or slightly modified reviews posted multiple times, possibly by bots or spam networks.

Challenges in Opinion Spam Detection

- Lack of labeled data: Genuine and spam reviews are difficult to distinguish without verified datasets.
- Language variation: Spammers use diverse phrasing to evade detection.
- User behavior ambiguity: Genuine reviewers might also write in exaggerated tones, making classification difficult.
- Evolving spam strategies: Spammers continuously adapt their methods to bypass detection algorithms.

Detection Techniques

1. Text-Based Methods
    - Keyword Analysis: Identifies spam using excessive use of promotional terms, URLs, or sentiment polarity.

    - Natural Language Processing (NLP): Utilizes tokenization, POS tagging, and syntactic structures to analyze review credibility.
    - TF-IDF & Word Embeddings: Transform textual content into numerical form for classification using machine learning.
2. Behavioral Analysis
    - Reviewer Behavior: Analyzes user's review patterns, such as frequency of reviews, similarity, and review timestamps.
    - Review Metadata: Includes user rating consistency, verified purchase status, and review helpfulness votes.
3. Machine Learning Approaches
    - Supervised Learning: Uses labeled datasets to train models such as Naïve Bayes, SVM, Random Forest, and Logistic Regression.
    - Unsupervised Learning: Clustering and anomaly detection without labeled data.
    - Deep Learning: Employs LSTM, CNNs, or transformer-based models (like BERT) to capture deeper textual patterns.

Dataset Sources

- Amazon Product Review Dataset
- Yelp Review Dataset
- Ott et al. Deceptive Opinion Spam Corpus (manually labeled fake/genuine hotel reviews)
- TripAdvisor and IMDB Reviews: Used for domain-specific opinion spam analysis.

Evaluation Metrics

- Accuracy: Overall correctness of the model.
- Precision and Recall: To balance false positives and false negatives.
- F1-Score: Harmonic mean of precision and recall.
- Confusion Matrix: Visual representation of predicted vs actual classes.

Applications

- E-commerce Platforms: Amazon, Flipkart, and eBay use spam detection to uphold trust.
- Review Aggregators: TripAdvisor and Yelp filter suspicious reviews.
- Recommendation Systems: Improve accuracy by excluding fake opinions.

CODE AND OUTPUT:

```
!pip install scikit-learn pandas numpy
```

```
import pandas as pd

df = pd.read_csv("opinion_reviews.csv")
print("Sample data:")
print(df.head())
print("\nClass distribution:\n", df['label'].value_counts())
```

```
Sample data:
                                              review  label
0            OMG! You need to buy this immediately!      0
1             Exactly what I needed. Great product.      1
2        Good value for money. Would purchase again.      1
3  I really liked this product, it works as expec...      1
4              Click here to win a free iPhone now!      0

Class distribution:
 label
0    200
1    200
Name: count, dtype: int64
```

```
import re

def clean_text(text):
    text = re.sub(r"http\S+", "", text)  # remove URLs
    text = re.sub(r"[^a-zA-Z]", " ", text)  # remove non-alphabet characters
    text = text.lower()
    return text

df['cleaned_review'] = df['review'].astype(str).apply(clean_text)

from sklearn.model_selection import train_test_split

X = df['cleaned_review']
y = df['label']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

from sklearn.feature_extraction.text import TfidfVectorizer

vectorizer = TfidfVectorizer(max_features=5000)
X_train_vec = vectorizer.fit_transform(X_train)
X_test_vec = vectorizer.transform(X_test)

from sklearn.linear_model import LogisticRegression

model = LogisticRegression()
model.fit(X_train_vec, y_train)
```

from sklearn.metrics import classification_report, confusion_matrix

y_pred = model.predict(X_test_vec)

print("Confusion Matrix:\n", confusion_matrix(y_test, y_pred))
print("\nClassification Report:\n", classification_report(y_test, y_pred))

```
Confusion Matrix:
 [[42  0]
 [ 0 38]]

Classification Report:
               precision    recall  f1-score   support

           0       1.00      1.00      1.00        42
           1       1.00      1.00      1.00        38

    accuracy                           1.00        80
   macro avg       1.00      1.00      1.00        80
weighted avg       1.00      1.00      1.00        80
```

CONCLUSION:

Opinion spam detection is a dynamic and essential area in the domain of natural language processing and machine learning. As online content continues to grow, robust and adaptable spam detection systems are vital. The fusion of NLP with behavioral analytics and advanced models like transformers holds great promise for future advancements in opinion spam detection.