Fake News Detection

Unmasking Deceptive Information

by Data Crusaders

Introduction

Fake news detection employs techniques such as natural language processing and machine learning to identify deceptive information presented as news.

Fake news can have serious consequences, from influencing elections to causing panic and confusion during times of crisis. In this presentation, we will explore the ways in which fake news spreads, the impact it can have, and most importantly, how we can detect it

Exploratory Data Analysis

Exploratory data analysis (EDA) plays a crucial role in effectively detecting fake news by providing insights into the characteristics that differentiate real and fake news articles.

During our EDA process, we examined various features, including **article length**, **unique word usage**, and **sentiment analysis**. The findings revealed that fake news articles tend to be shorter in length and employ emotionally charged language, while real news articles exhibit more formal writing styles. Furthermore, fake news articles often contain a higher number of spelling and grammatical errors compared to their genuine counterparts. By leveraging these insights, we can develop more accurate and reliable fake news detection models.

Metric and Model Selection

Logistic Regression

Simple and interpretable, performs well with small to moderate-sized datasets, provides probabilistic interpretations.
(Accuracy: 0.98)

Decision Tree

Simple and interpretable, handles numerical and categorical features, captures non-linear relationships
(Accuracy: 0.99)

Random Forest

Ensemble learning, handles high-dimensional data well, captures complex relationships.
(Accuracy: 0.99)

Accuracy Caparision

- Among the three models, the Decision Tree algorithm achieved the highest accuracy.
- This indicates that the Decision Tree model was able to capture the underlying patterns in the data effectively.
- Using Intel extensions in scikit-learn, the models achieved higher accuracy in fake news detection due to improved performance and optimized computations.

Model Implementation

Decision Tree:

from sklearn.tree import DecisionTreeClassifier patch_sklearn(
extension for Scikit-learn time: {train_patched:.2f} s"

```
unpatch_sklearn()
start = timer()
DT = DecisionTreeClassifier()
DT.fit(X_train, Y_train)
train_patched = timer() - start
f"Scikit-learn time: {train_patched:.2f} s"
```

```
'Scikit-learn time: 22.84 s'
```

t

```
pred_dt = DT.predict(X_test)
DT.score(X_test, Y_test)
```

```
0.9966592427616926
```

```
print(classification_report(Y_test, pred_dt))
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	4696
1	1.00	1.00	1.00	4284
accuracy			1.00	8980
macro avg	1.00	1.00	1.00	8980
weighted avg	1.00	1.00	1.00	8980

Intel extensions provided optimized implementations of the Logistic Regression, Decision Tree and Random Forest models, resulting in improved performance and efficiency during training and prediction, leading to faster execution and enhanced accuracy in fake news detection.

Future Works

- Develop advanced machine learning models, including **deep learning architectures.**
- Integrate real-time news feeds and social media monitoring for timely identification of emerging fake news.
- Explore multi-modal approaches, combining text, image, and video analysis to detect deceptive content across various media formats.

Reference

- Pandas Documentations
- Logistic Regression Documentations
- Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2019). Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 21(2), 22-36.

Conclusion

In summary, fake news detection is a vital field where advanced machine learning models, real-time monitoring, and multi-modal analysis are being employed to enhance accuracy and timely identification of deceptive information. By leveraging these approaches, we can mitigate the impact of fake news and promote a more trustworthy information ecosystem.

THANK YOU!