# **Decision Tree and Random Forest Evaluation Report**

This section of the report presents a comprehensive analysis of the Decision Tree and Random Forest classifiers used in our sentiment classification project. We evaluate their performance using both Bag-of-Words (BoW) and TF-IDF features and analyze the effect of key hyperparameters on model performance.

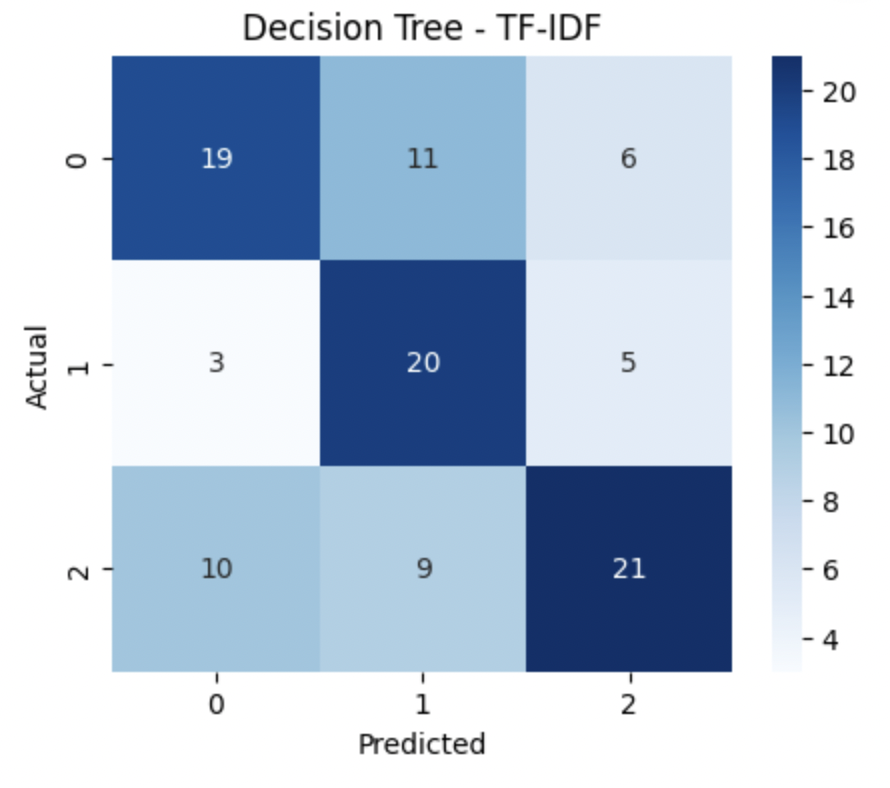
## **1. Decision Tree with TF-IDF Features**

**Accuracy :** 57.69%

**Classification Report:**



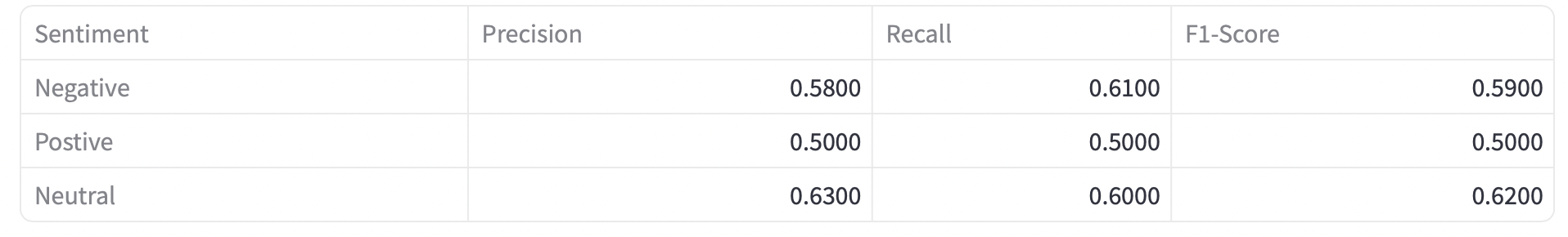
**Confusion Matrix:**



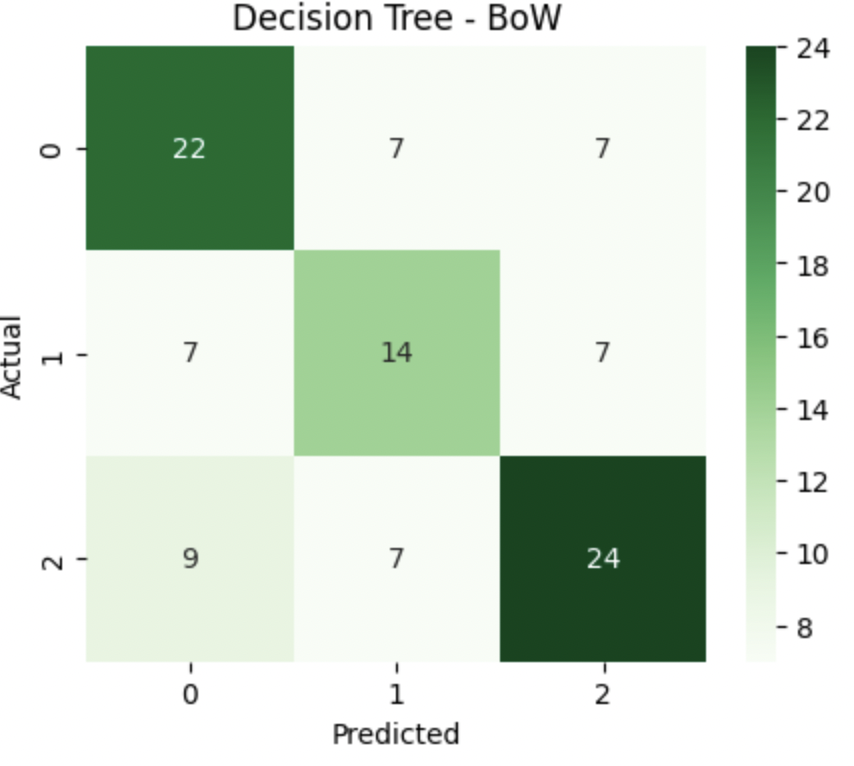
## **2. Decision Tree with BoW Features**

**Accuracy :**57.69%

**Classification Report:**



**Confusion Matrix:**



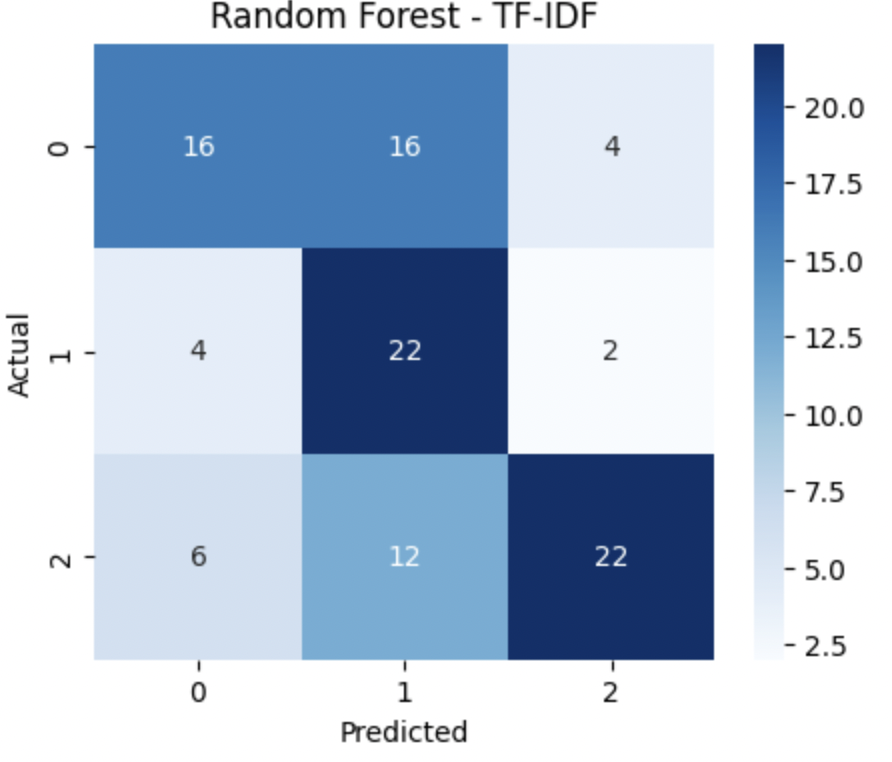
## **3. Random Forest with TF-IDF Features**

**Accuracy :**57.69%

**Classification Report:**



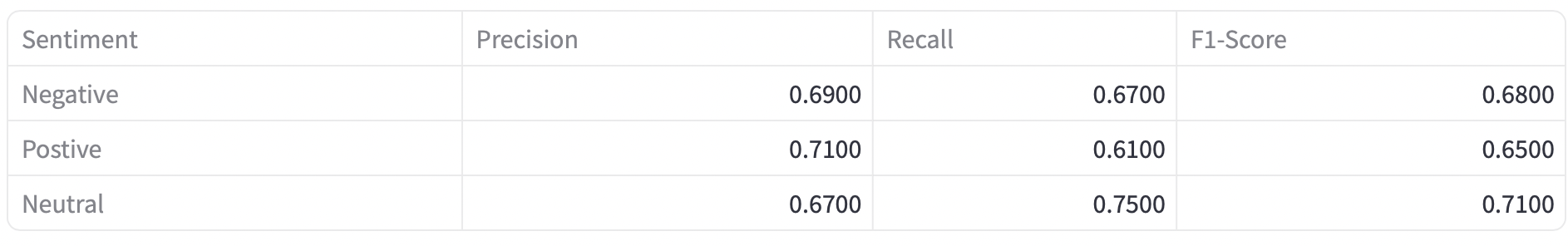
**Confusion Matrix:**



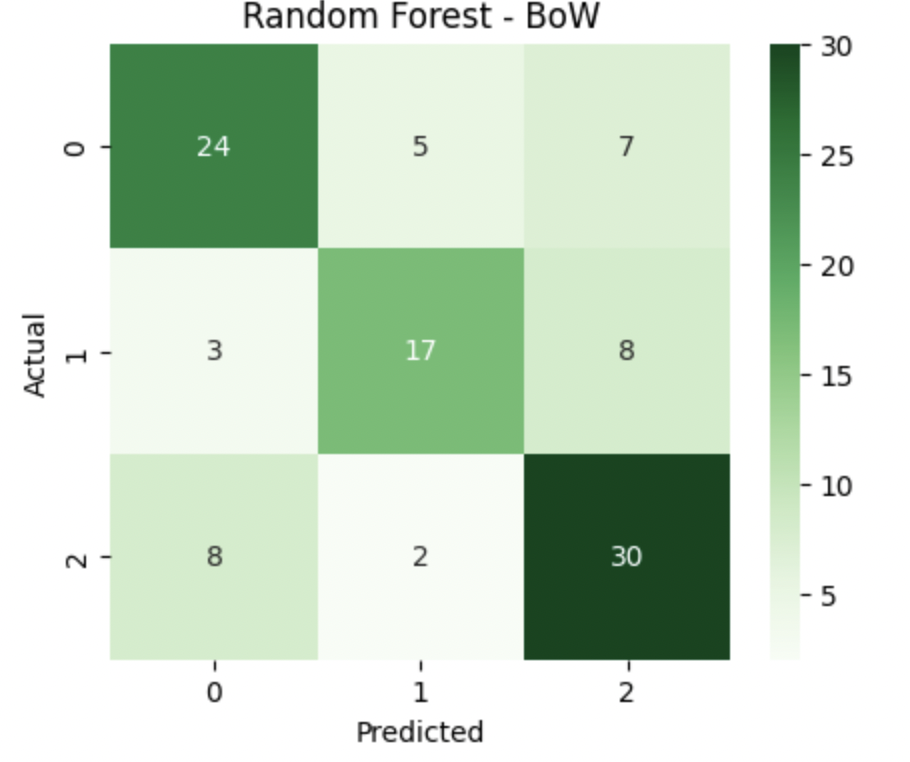
## **4. Random Forest with BoW Features**

**Accuracy :** 68.27%

**Classification Report:**



**Confusion Matrix:**



## **Why Random Forest with BoW Outperforms TF-IDF**

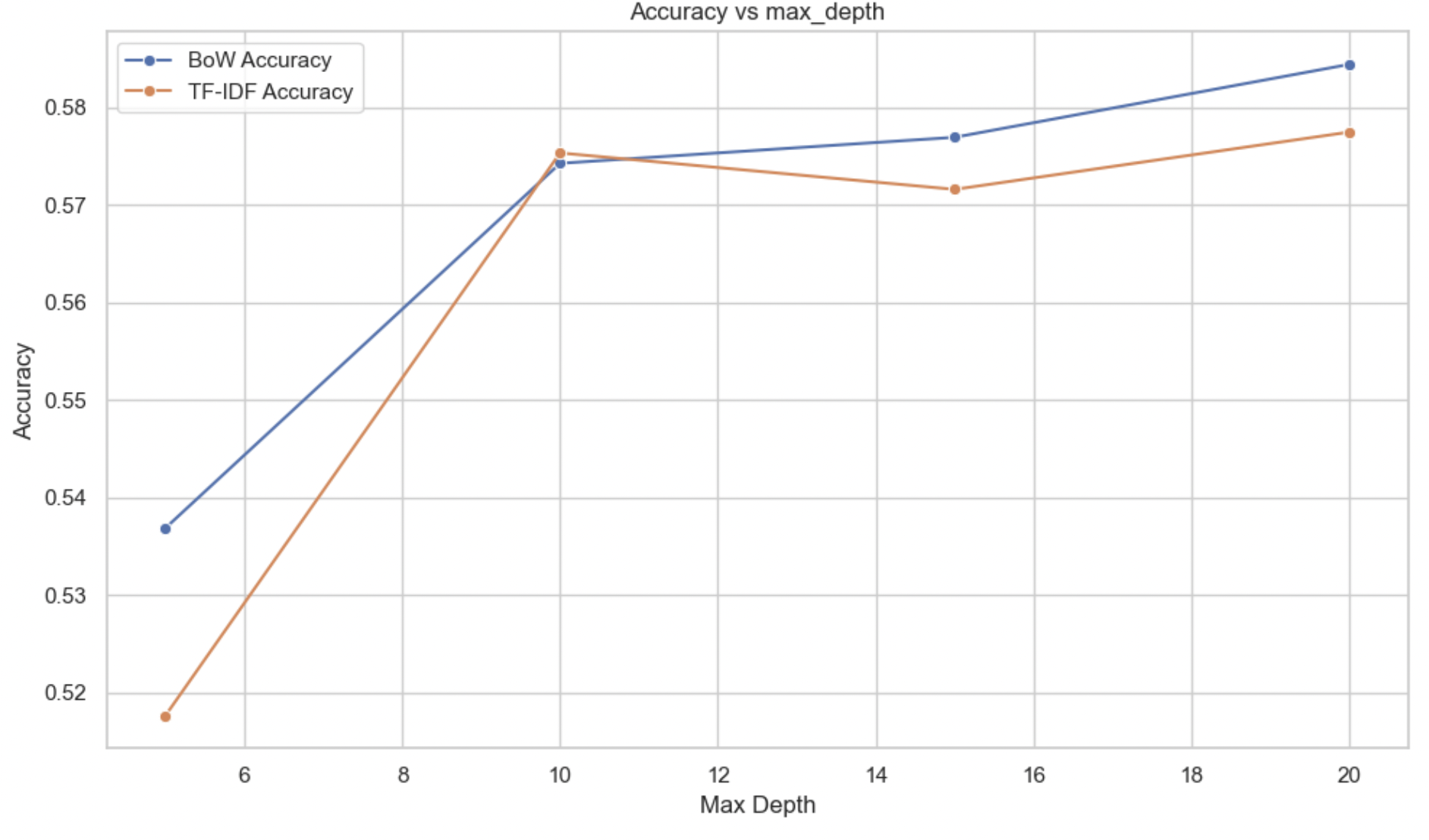
From our model evaluations, it is evident that Bag-of-Words (BoW) features outperform TF-IDF features in this sentiment classification task, particularly when paired with a Random Forest classifier. This may be attributed to the nature of our dataset — where common sentiment-indicative words (e.g., *good*, *bad*, *love*, *hate*) are crucial for classification. TF-IDF tends to down-weight these frequent terms, which can reduce their impact on model performance.

On the other hand, Random Forest classifiers consistently outperform Decision Trees due to their ensemble learning approach. By aggregating predictions from multiple decision trees, Random Forest reduces overfitting and improves generalization.

Thus, the best overall performance was achieved using BoW features with a Random Forest classifier, making it the recommended setup for this sentiment analysis task based on our current dataset.

## **Hyperparameter Analysis**

### **Accuracy vs. Max Depth**

****

**Interpretation:**

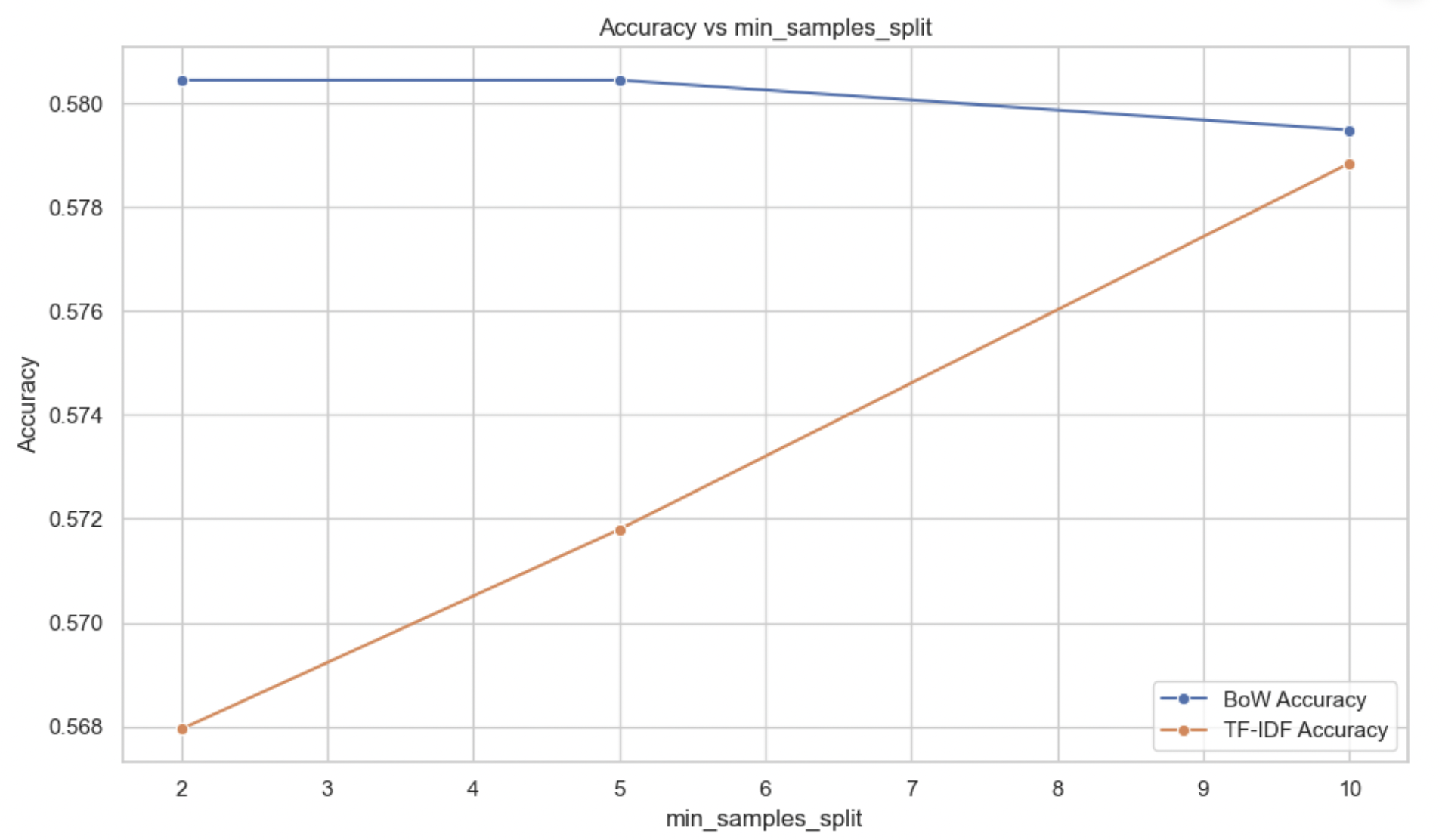
* Accuracy increases with depth for both BoW and TF-IDF initially.
* BoW consistently outperforms TF-IDF across all depth values.
* Beyond a depth of 10, both lines plateau, with BoW still slightly ahead.

**Explanation:**

* Shallow depths underfit the data, especially for TF-IDF.
* Optimal performance is reached around depth 10.
* Deeper trees risk overfitting, especially with TF-IDF.

**Conclusion:** A max depth between 10–15 is ideal, with BoW maintaining better generalization.

### **Accuracy vs. Min Samples Split**

****

**Interpretation:**

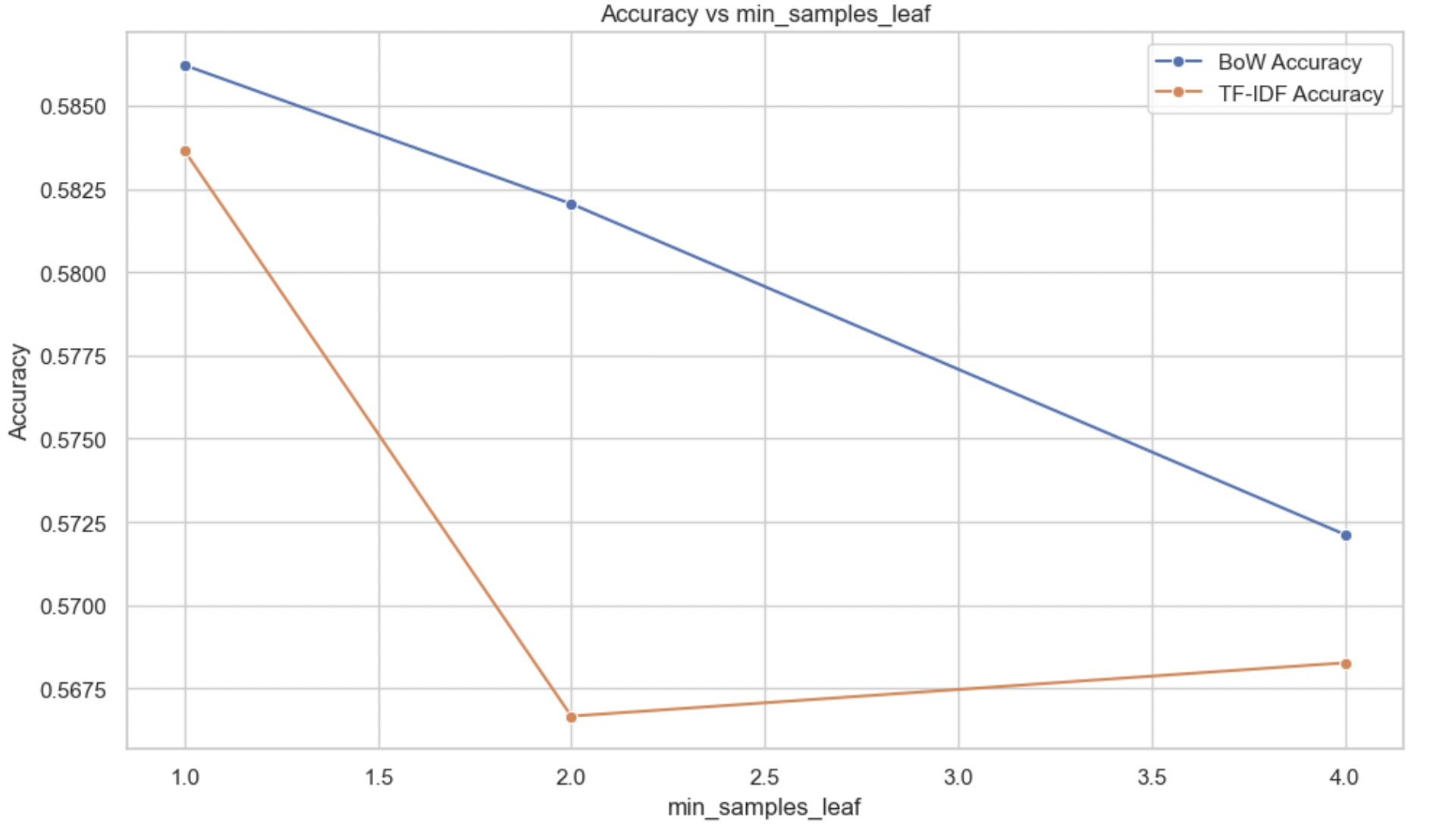
* BoW starts with higher accuracy and remains stable.
* TF-IDF starts low but improves with increased min\_samples\_split.

**Explanation:**

* Small splits increase model complexity and risk overfitting.
* BoW performs well at small splits; TF-IDF improves with regularization.

**Conclusion:** TF-IDF benefits from higher min\_samples\_split (~10) to reduce overfitting. BoW is robust even at lower values.

### **Accuracy vs. Min Samples Leaf**

****

**Interpretation:**

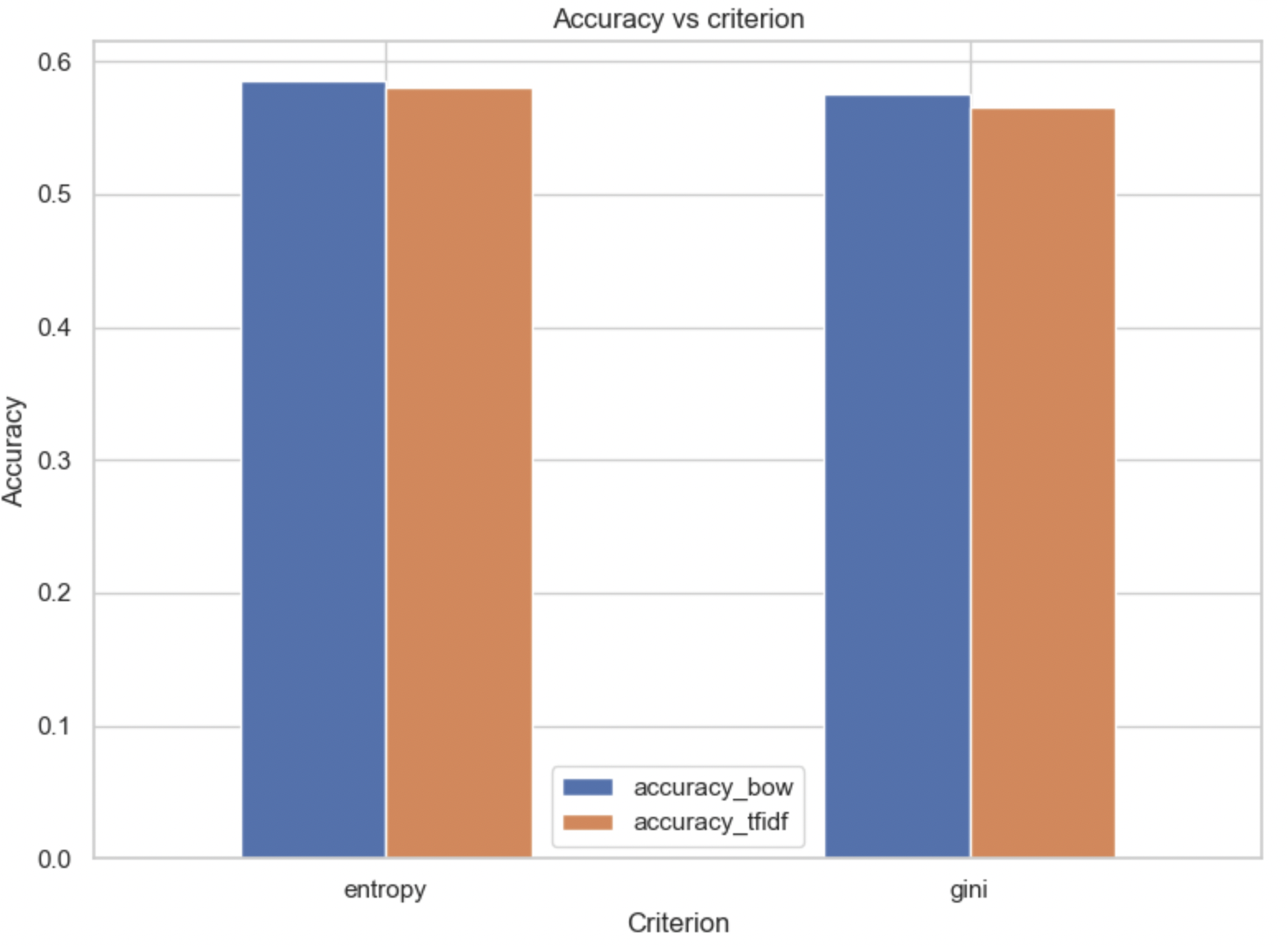
* BoW peaks at min\_samples\_leaf = 1 and declines steadily.
* TF-IDF shows a sharp drop at 2, then stabilizes at a lower level.

**Explanation:**

* BoW thrives with minimal leaf constraints due to frequent, strong features.
* TF-IDF over-regularizes easily, losing performance.

**Conclusion:** Keep min\_samples\_leaf low for BoW; apply slight regularization for TF-IDF.

### **Accuracy vs. Criterion (Entropy vs. Gini)**

****

| **Criterion** | **BoW Accuracy** | **TF-IDF Accuracy** |
| --- | --- | --- |
| Entropy | ~0.586 | ~0.581 |
| Gini | ~0.578 | ~0.567 |

**Interpretation:**

* Entropy consistently outperforms Gini across both feature sets.
* BoW outperforms TF-IDF for both criteria.

**Explanation:**

* Entropy captures subtle splits in sparse data better.
* Gini is faster but slightly less precise.

**Conclusion:** Entropy is the preferred criterion, especially with BoW, for capturing key sentiment-indicative patterns.

## **Final Summary**

* BoW features paired with Random Forest yield the highest accuracy.
* TF-IDF struggles due to its down-weighting of high-frequency, sentiment-rich words.
* Hyperparameter tuning reveals that moderate tree depth and slight regularization (only for TF-IDF) improve generalization.
* BoW remains more robust to parameter changes and consistently delivers strong performance.

**Recommendation:** Use Random Forest with BoW, max\_depth = 10–15, and min\_samples\_split = 2–5 for optimal results.