**Text Analysis Report**

**1. Introduction & Objective**

This project analyzes a set of text documents to extract meaningful insights using **Natural Language Processing (NLP)** techniques.

**Objectives:**

* Identify the most important words in each document using **TF-IDF**.
* Discover semantic relationships between words using **Word2Vec**.
* Extract hidden topics across the corpus using **LDA Topic Modeling**.

**2. Methodology**

The following steps were followed for the analysis:

1. **Data Preprocessing**:
   * Converted text to lowercase.
   * Removed punctuation, numbers, and stopwords.
   * Tokenized the text into individual words.
2. **TF-IDF Analysis**:
   * Calculated TF-IDF scores for all words.
   * Identified the top words per document to highlight important terms.
3. **Word2Vec Analysis**:
   * Trained a Word2Vec model to capture semantic relationships between words.
   * Found words similar to key terms (e.g., phone, movie, food).
4. **LDA Topic Modeling**:
   * Applied Latent Dirichlet Allocation (LDA) to uncover hidden topics across documents.
   * Identified clusters of related words representing dominant themes.

**3. Results & Observations**

**TF-IDF Results:**  
The top words per document reveal document-specific important words.

* Document 1: phone, battery, camera
* Document 3: movie, action, exciting
* Document 4: food, restaurant, delicious

**Word2Vec Results:**  
Captures semantic similarities between words.

* phone → camera, smartwatch, laptop
* movie → concert, action, plot
* food → restaurant, delicious, meal

**LDA Topic Modeling Results:**  
Reveals hidden topics in the corpus.

* Topic 0: battery, phone, laptop, camera
* Topic 1: movie, action, concert, plot
* Topic 2: food, restaurant, delicious
* Topic 3: software, game, graphics
* Topic 4: travel, experience, scenery

**4. Discussion & Conclusion**

Combining TF-IDF, Word2Vec, and LDA provides a comprehensive understanding of the text corpus.

**Key Takeaways:**

* TF-IDF identifies document-specific keywords.
* Word2Vec captures semantic relationships between words.
* LDA uncovers main topics across documents.
* Dominant themes in the dataset include **Electronics, Entertainment, Food, and Travel**.

This analysis highlights the importance of preprocessing, vector representation, and topic modeling for understanding text data effectively.