**Black Friday Sales Prediction - Task 2 Final Report**

**1. Project Overview**

**Objective:**

This project applies **text mining techniques** to predict Black Friday sales by integrating structured sales data with unstructured customer reviews. The goal is to extract valuable insights from text data to enhance purchase amount predictions.

**Methods Used:**

1. **Structured Data Processing** – Sales data (gender, age, product categories, city demographics, purchase amounts).
2. **Text Mining & Sentiment Analysis** – Extracted sentiment polarity from customer reviews.
3. **Feature Engineering** – Merged sentiment scores with structured sales data.
4. **Machine Learning Models** – Trained and evaluated multiple models for sales prediction.

**2. Data Analysis & Key Findings**

**Insights from Structured Data:**

* **Male customers** contribute to 75% of total purchases.
* The **26-35 age group** spends the most.
* **Unmarried customers** are more likely to make higher purchases.
* **City B** has the most purchases, but **City C** customers spend more per person.

**Insights from Text Mining:**

* **Positive sentiment** in reviews correlates with higher purchase amounts.
* **Negative sentiment** is linked to lower spending.
* Sentiment analysis helps identify **customer satisfaction trends**, useful for promotions and product recommendations.

**3. Model Performance & Evaluation**

We trained multiple machine learning models:

|  |  |
| --- | --- |
| **Model** | **RMSE Score** |
| **Linear Regression** | High Error (Not Optimal) |
| **Decision Tree Regressor** | Moderate |
| **Random Forest Regressor** | Good |
| **XGBoost Regressor** | **Best Model - RMSE: 2879 - 2957** |
|  |  |

**4. Business Recommendations**

🔹 **Target High-Spending Demographics** – Personalize offers for **males, age 26-35, and unmarried customers**. 🔹 **Optimize Pricing for City C Customers** – Since they spend more per person, focus premium promotions here. 🔹 **Leverage Sentiment Trends** – Improve customer experience by addressing **negative sentiment trends**. 🔹 **Boost Positive Engagement** – Encourage satisfied customers to leave reviews, improving brand reputation.

**5. Conclusion**

By integrating structured sales data with **text mining techniques**, we improved sales predictions. **XGBoost performed best with an RMSE of ~2879**, showing the potential of **text data in predicting purchasing behavior**. Future improvements can include **more review data, topic modeling, and hyperparameter tuning**.