Machine Learning Internship Assignment Report

**Assignment**: Predictive Model for Digital Marketing Campaign Performance

**Introduction**

The objective of this assignment was to conduct an Exploratory Data Analysis (EDA) and to develop a predictive model that estimates the performance of a new digital marketing campaign based on its creative attributes. The model focuses on predicting conversions using various features related to the campaign's creative content.

**Methodology**

**Exploratory Data Analysis (EDA)**

We conducted an in-depth analysis of various creative attributes to understand their impact on key performance metrics in digital marketing campaigns. Interactive plots were generated using Plotly to visually explore relationships between creative attributes and performance metrics.

For **performance metrics**, a bar chart is used to show the relationship between the given performance metrics and time.

**Spend vs Time**: Each bar will represent the spending for a specific date.

**Impression vs Time:** Each bar will represent the impression for a specific date.

**Like count vs Time:** Each bar corresponds to a specific date and its height represents the number of likes.

**Comment count vs Time:** Each bar corresponds to a specific date and its heights represent the number of comments.

**Repost count vs Time:** Each bar corresponds to a specific date and its height represents the number of reposts.

**Total Engagement vs Time:** Each bar corresponds to a specific date and its heights represent the total engagement (likes, comments and repost combined).

For **creative attributes**, different plots are used for Exploratory Data Analysis (EDA).

**Text Area % vs Conversion**: This scatter plot examines the relationship between the percentage of the creative area occupied by text and the conversions.

**Video URL vs Conversion**: This histogram visualizes the presence or absence of a video URL in the creative content and how it relates to conversions. It provides insights into whether including a video URL affects conversion rates.

**Dominant Color:** This pie chart visualizes the distribution of dominant colors in the creative content. It provides insights into the color composition and its potential impact on conversions.

**Logo vs Conversion:** This histogram visualizes the presence or absence of logos in the creative content and how it relates to conversions. It provides insights into whether including logos affects conversion rates.

**Performance Metrics vs Creative Attributes**

Performance Metric: **Total Engagements** vs. Creative Attribute: **Dominant Color**

The box plot compares the dominant color in the creative content with the total number of engagements.

Performance Metric: **Conversion** vs. Creative Attribute: **Logo Presence**

This histogram visualizes the presence or absence of logos in the creative content and how it relates to conversions.

Performance Metric: **Comment Count** vs. Creative Attribute: **Text Area Percentage**

This scatter plot compares the number of comments with the percentage of the creative area occupied by text.

**Machine Learning Model**

**Data Collection and Preprocessing**

The dataset consisted of information on past digital marketing campaigns, including attributes related to creative content such as visual elements, messaging, and format. The data underwent a rigorous preprocessing pipeline, which included handling missing values, encoding categorical variables, and scaling numerical features.

**Feature Selection and Engineering**

To ensure the model's efficiency and interpretability, a feature selection process was employed. Using techniques such as SelectKBest with mutual\_info\_regression, the most relevant attributes were identified.

**Model Selection and Tuning**

A regression approach was chosen to predict conversion rates. Several regression models, including Linear Regression, Random Forest Regressor, and Decision Tree Regressor, were considered.

**Model Evaluation**

The model's performance was assessed using both training and testing sets. The results are encouraging, with a training accuracy of 92% and a testing accuracy of 78%.

**Conclusion**

**Exploratory Data Analysis (EDA)**

The analysis of creative attributes in digital marketing campaigns has provided valuable insights into their impact on key performance metrics. Through the exploration of various attributes such as post content length, presence of faces, text area percentage, dominant color, and others, we gained a deeper understanding of their influence on conversions, engagements, likes, and comments.

In conclusion, a thoughtful approach to creative attribute selection and optimization can have a substantial impact on the success of digital marketing campaigns. By leveraging these insights, marketers can refine their content strategies to enhance user engagement, increase conversions, and ultimately drive better campaign performance.

**Machine Learning Model**

The developed predictive model provides valuable insights into estimating the performance of new digital marketing campaigns based on their creative attributes. The high training accuracy indicates that the model effectively learns from the data.

**Recommendations**

**Exploratory Data Analysis (EDA)**

Based on our analysis, we offer the following recommendations for optimizing digital marketing campaigns:

1. Consider creating longer engaging post content to potentially increase conversion rates.
2. Incorporate faces strategically in creative content to enhance user engagement and like counts.
3. Avoid excessive text content to maintain optimal conversion rates.

**Machine Learning Model**

Further experimentation with different regression algorithms and feature engineering techniques could lead to improvements in model performance. Additionally, incorporating more data from diverse campaigns may enhance the model's generalization capabilities.