ASWIN.G.KUMAR

I have Completed MA English and Currently I am Pursuing Master Certificate Program in Artificial Intelligence

Toc H International Academy

SREE SANKARACHARYA COMPUTER CENTRE

Since 1995

SREE SANKARACHARYA COMPUTER CENTRE PVT. LTD.

PRAGYAN AI

Grow with Gyan

Instagram Reach Analysis & Forecasting

1

Using Machine Learning Both Regression & Classification Models

Introduction

ın this analysis, we aim to forecast Instagram reach over time based on temporal features such as Day, Month, and Day of the Week. The goal is to predict how Instagram reach evolves across different time periods using regression and classification models. Instagram reach represents the total number of unique users who have seen a specific post or content over a certain period, which is crucial for social media strategies and audience engagement analysis.

We utilize several regression and classification techniques to model Instagram reach: Random Forest Regressor, Decision Tree Regressor, and Gradient Boosting Regressor, Random Forest Classifier, Decision Tree Classifier and etc. These models are chosen due to their ability to handle non-linear relationships and capture complex patterns in the data.

Social media platforms like Instagram play a crucial role in digital marketing and brand awareness. Analyzing and forecasting Instagram reach helps businesses and influencers optimize their content strategies to maximize engagement. This project leverages machine learning models to predict Instagram reach using temporal features like day, month, and day of the week. Additionally, classification models categorize Instagram reach into different levels to provide further insights

Objective

- Analyze historical Instagram reach data.
- Identify patterns and trends in reach over time.
- Predict future Instagram reach using regression models.
- Categorize Instagram reach levels using classification models.
- Compare different machine learning models to determine the best performer

Advantages of the Project

1. **Data-Driven Decision Making**: Helps businesses and influencers optimize content posting strategies.
2. **Improved Engagement**: Predicting reach trends enables better planning for peak engagement.
3. **Model Comparison:** Evaluates multiple regression and classification models to determine the most accurate prediction method.
4. **Feature Extraction**: Uses temporal features (day, month, weekday) to enhance model performance.
5. **Automation**: Reduces manual effort in analyzing social media performance.
6. **Visualization:** Uses various plots and graphs to present insights effectively.

Disadvantages of the Project

1. **Limited Features:** The model only considers temporal features and does not account for external factors like hashtags, content type, or engagement metrics.
2. **Data Dependency:** Performance is highly dependent on the quality and quantity of the dataset.
3. **Model Complexity:** Some models (e.g., Gradient Boosting) may require more computational power and fine-tuning.
4. **Short-Term Forecasting:** The model may struggle with long-term predictions due to changing social media trends.

Methodology

1. Data Preprocessing

- Loaded Instagram reach dataset.
- Converted the date column to a datetime format.
- Extracted temporal features: day, month, and day of the week.
- Checked for missing values and performed necessary cleaning.

2. Exploratory Data Analysis (EDA)

- Visualized Instagram reach trends over time.
- Used correlation heatmaps to analyze feature relationships.
- Bar plots to observe distribution patterns.

3. Regression Models for Reach Prediction

- **Random Forest Regressor**
- **Decision Tree Regressor**
- **Gradient Boosting Regressor**
- Models were evaluated using **Mean Squared Error (MSE) and R^2 Score**.

4. Classification Models for Reach Categorization

- Defined four reach categories: Low, Medium, High, and Very High.
- **Random Forest Classifier**
- **Decision Tree Classifier**
- **Gradient Boosting Classifier**
- Models were evaluated using **Accuracy, Precision, Recall, and F1 Score**.

5. Model Comparison

- The best regression model was selected based on the lowest **MSE** and highest **R^2 Score**.
- The best classification model was selected based on **Accuracy and F1 Score**.

Conclusion

In this study, we analyzed Instagram reach data and developed predictive models for both regression and classification tasks. For regression, **Random Forest Regressor** provided the best results with the highest R-squared value and the lowest Mean Squared Error (MSE), making it the most effective model for predicting future Instagram reach.

For classification, the **Decision Tree Classifier** emerged as the best-performing model with the highest **accuracy (76.71%)**, **precision (79.99%)**, and **F1 score (78.21%)**, indicating its superior ability to categorize Instagram reach into different levels.

Our findings suggest that **Decision Tree Classifier** is the most reliable for classifying Instagram reach levels, while **Random Forest Regressor** is the best for forecasting future engagement. These models can help social media analysts optimize content strategies and improve reach by leveraging predictive analytics.