D. Saul Jameson
https://www.linkedin.com/in/dsauljameson/

Project Documentation: Predicting Conversion Rates & SEO Performance Using Gradient Boosting

Project Overview

This project focuses on using **Gradient Boosting Regression (XGBoost)** to predict **conversion rates** based on **SEO factors**. The model helps businesses optimize **search engine marketing (SEM)** by identifying the impact of **keywords, bid strategies, time of day, and device types** on conversions. The predictions are visualized using **Plotly** and presented in an **interactive HTML dashboard**.

Use Case: SEO & Conversion Rate Optimization

Businesses investing in **paid search advertising** need to know:

- **Which keywords drive the most conversions**
- **What time of day delivers the best performance**
- **How device type affects conversion rates**
- **How competitor activity influences ad effectiveness**

By leveraging **machine learning**, the model helps marketers **allocate ad budgets more effectively**, improve **bidding strategies**, and optimize **campaign timing** for **higher ROI**.

Data Collection & Preprocessing

Data Sources

- The dataset for this project was **synthetically generated** to simulate realistic SEO and ad performance data. The data includes variables such as keyword type, time of day, bid amounts, device type, and competitor density, ensuring a diverse range of scenarios for model training and evaluation.

Preprocessing Steps

- **Datetime Features**: Extracted time of day and day of the week for conversion trends.
- **Categorical Encoding**: Converted keyword types and device types into numerical format.
- **Feature Scaling**: Standardized numerical features to improve model performance.

Packages Used

```
import pandas as pd
import numpy as np
import plotly.express as px
import xgboost as xgb
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler, OneHotEncoder
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
```

Exploratory Data Analysis (EDA)

Plotly visualizations were used to explore the dataset and identify key patterns.

Conversion Rate Over Time

A **line chart** was used to detect seasonal trends.

```
fig = px.line(df, x="time_of_day", y="conversion_rate", title="Conversion Rate by Time of Day")
fig.show()
```

Feature Importance in Predicting Conversions

XGBoost's built-in feature importance was plotted to see which factors influenced conversion rates the most.

```
fig = px.bar(importance_df, x="Importance", y="Feature", orientation="h",
             title="Feature Importance in Predicting Conversion Rate",
             color="Importance", color_continuous_scale="Blues")
fig.show()
```

D. Saul Jameson
https://www.linkedin.com/in/dsauljameson/

Modeling Approach

1. Data Splitting

The dataset was split into **80% training, 20% testing**.

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

2. XGBoost Model Training

The **Gradient Boosting model** was used for regression to predict **conversion rates**.

```
xgb_model = xgb.XGBRegressor(objective="reg:squarederror", n_estimators=500,
learning_rate=0.05, max_depth=6)
xgb_model.fit(X_train, y_train)
```

3. Model Evaluation

The model's performance was measured using **R², MAE, and RMSE**.

```
y_pred = xgb_model.predict(X_test)
r2 = r2_score(y_test, y_pred)
mae = mean_absolute_error(y_test, y_pred)
rmse = np.sqrt(mean_squared_error(y_test, y_pred))

print(f"R²: {r2:.4f}, MAE: {mae:.4f}, RMSE: {rmse:.4f}")
```

Metric	Value	Interpretation
R² Score	**0.9986**	Model explains 99.86% of variance (very high).
MAE	**0.0021**	Average prediction error is very low.
RMSE	**0.0030**	Errors are minimal, indicating strong predictive performance.

Feature Importance Analysis

Rank	Feature	Importance	Meaning

D. Saul Jameson
https://www.linkedin.com/in/dsauljameson/

1	**Time of Day**	**0.362**	The strongest predictor of conversions. Ads perform best at specific times.
2	**Informational Keywords**	**0.190**	Informational searches drive conversions more than expected.
3	**Competitor Keywords**	**0.174**	Users searching for competitor terms convert at a high rate.
4	**Mobile Device**	**0.139**	Mobile users convert more than tablet/desktop users.
5	**Transactional Keywords**	**0.064**	Less influential than expected.
6	**Bid Amount, CPC, CTR**	**Low**	Spending more does not guarantee better conversions.

Final Visualization & HTML Dashboard

An **interactive dashboard** was created using **Plotly and HTML**.

```
from plotly.io import to_html

dashboard_html = f"""
<!DOCTYPE html>
<html lang="en">
<head>
   <meta charset="UTF-8">
   <title>Conversion Rate Dashboard</title>
</head>
<body>
   <h1 style="text-align: center;">Conversion Rate Dashboard</h1>
   <div style="display: flex; flex-wrap: wrap; justify-content: space-around;">
      <div>{to_html(fig1, full_html=False)}</div>
      <div>{to_html(fig2, full_html=False)}</div>
      <div>{to_html(fig3, full_html=False)}</div>
   </div>
</body>
</html>
"""

with open("Conversion_Rate_Dashboard.html", "w") as f:
```

D. Saul Jameson
https://www.linkedin.com/in/dsauljameson/

```
    f.write(dashboard_html)

print("Dashboard saved as Conversion_Rate_Dashboard.html")
```

Key Takeaways

1. **Time of day is the most important factor in conversion rates.** Optimizing ad schedules is more effective than increasing CPC.

2. **Informational and competitor keywords drive more conversions than transactional ones.** This challenges the assumption that **users searching for direct transactions** convert best.

3. **Mobile users convert better than desktop and tablet users.** The ad experience should be optimized for mobile.

4. **CTR, CPC, and bid amount have minimal influence.** More spending does not necessarily mean better conversions.

Next Steps

- **A/B testing ad schedules** based on peak conversion times.
- **Refining keyword strategy** to prioritize informational and competitor search terms.
- **Further segmentation by industry or location** to improve targeting precision.
- **Deploying the dashboard as a live web app** for real-time monitoring.

Conclusion

This project successfully built a **Gradient Boosting regression model** to predict **conversion rates based on SEO performance**. The results provide actionable insights for **ad targeting, budget allocation, and search engine marketing strategy**. The model's insights were visualized using **Plotly** and presented in an **interactive HTML dashboard**.

D. Saul Jameson
https://www.linkedin.com/in/dsauljameson/