

ML
PROJECT # 1
PART # 2

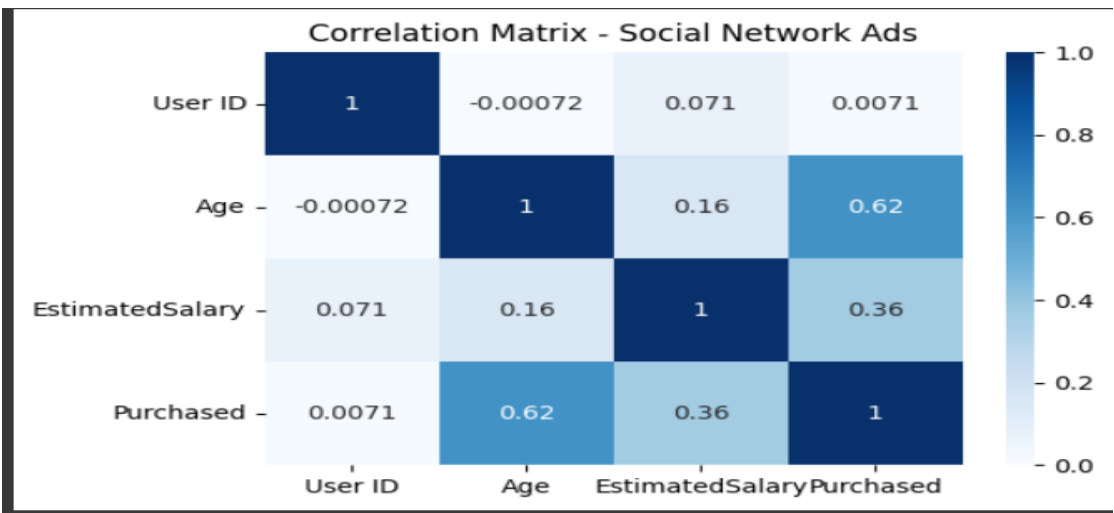
SHEHLA SHIREEN

Classification Report: Predicting Purchase Behavior

Dataset: Social_Network_Ads.csv

Objective: Predict whether a user will purchase a product based on Age, Estimated Salary, and Gender.

1. Data Insights



- Strongest Predictor:** Age (correlation with purchase = **0.62**)
- Moderate Predictor:** Estimated Salary (**0.36**)
- Weak Predictor:** Gender (near-zero correlation = **-0.04**)
- Conclusion:** Age and Salary are key predictors. Gender has minimal impact.

2. Model Performance Overview

Model	Accuracy	Precision	Recall	F1-Score	Confusion Matrix
Naive Bayes	0.93	0.94	0.865	0.901	[[61, 2], [5, 32]]

Model	Accuracy	Precision	Recall	F1-Score	Confusion Matrix
KNN (k=3)	0.91	0.868	0.892	0.880	[[58, 5], [4, 33]]
KNN (k=5)	0.92	0.872	0.919	0.895	[[58, 5], [3, 34]]
KNN (k=7)	0.93	0.875	0.946	0.909	[[58, 5], [2, 35]]
Decision Tree (Gini)	0.84	0.800	0.757	0.778	[[56, 7], [9, 28]]
Decision Tree (Entropy)	0.84	0.800	0.757	0.778	[[56, 7], [9, 28]]

Best Model: KNN (k=7) – highest balance of precision, recall, and F1-score.

Naive Bayes is a strong, simple alternative.

Decision Trees underperform due to likely overfitting and poor generalization.

3. Key Observations

- KNN (k=7) had only **2 false negatives** and **5 false positives**.
- Naive Bayes avoided more false positives but missed more actual purchasers.
- Models using continuous features (Age and Salary) performed better.
- Decision Trees may need tuning or pruning.

Conclusion

KNN (k=7) is the most reliable model for this dataset. Naive Bayes is a strong backup for simpler deployment. Focus on Age and Salary for predictive modeling; Gender adds little value.