Interview Mini Project- Instagram likes Regression or Classification

Data Extraction

Data out of (Q1 – factor * IQR, Q3 + factor * IQR) are defined as outliers. The factor is set to 5, because it can maintain only 5% data were dropped.

For Classification, to classify data into 3 levels, 33% of likes are Low level, 33% - 67% of likes are Medium level, and 67% - 100% of likes are High level.

Data Splitting

Traditional Model: Test: Train = 2 : 8 / Neural Network: Test: Validate: Train = 2: 2: 6

(I tried to make all models using same training data and test data.)

Model Selection

Random Forest, KNN, Gaussian NB, Multinomial NB

Neural Network (Vision Transformer)

Result

	Accuracy	Low Precision	Medium Precision	High Precision
Random Forest	0.77	0.83	0.68	0.81
KNN	0.55	0.54	0.51	0.62
Gaussian NB	0.3	0.33	0.11	0.32
Multinomial NB	0.59	0.57	0.5	0.78
Neural Network	0.32	0.31	0.33	0.33

Random Forest > KNN or Multinomial NB > Gaussian NB or Neural Network

In summary, the Random Forest model emerges as the most promising choice based on the performance metrics. It exhibits the highest precision across all categories of like levels. However, its precision in the "Medium" category is lower, which can be attributed to the inherent difficulty in distinguishing "Medium" likes, as they share similarities with both lower and higher categories. Additionally, result of different parameter compositions test indicates that "no_of_comments" and "follower_count_at_t" dominates the prediction accuracy.

On the other hand, both KNN and Multinomial Naive Bayes (NB) models yield similar accuracies, but they differ in precision among the various like levels. KNN's precision remains relatively consistent across categories, while Multinomial NB excels in predicting "High" level likes.

Finally, the Vision Transformer model lags behind with the lowest accuracy and precision. Across all three like level categories, the precision approximates the average ratio of 1/3. This indicates that the model struggles to use image content as a reliable classifier for predicting like levels.