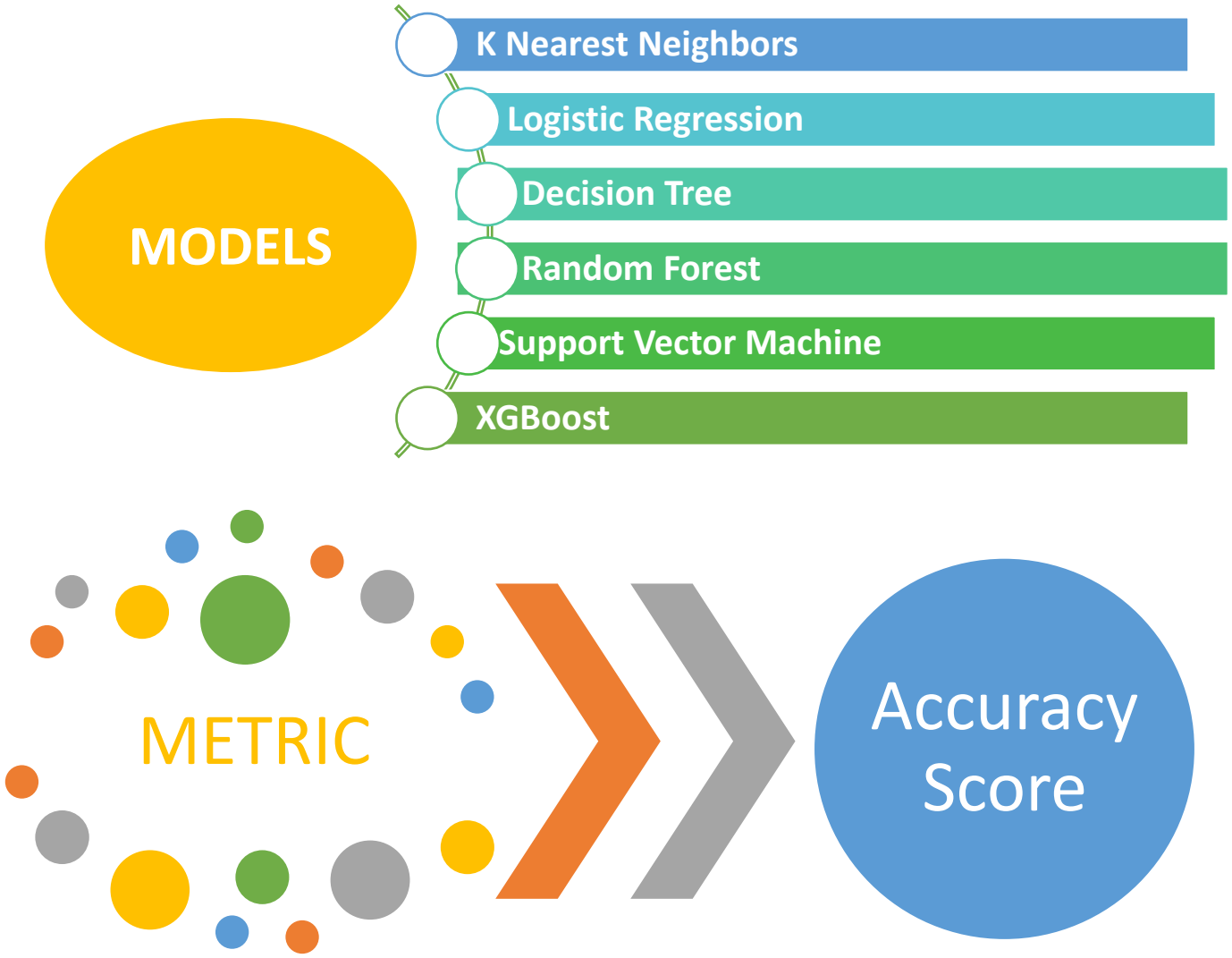
Data Scientist Professional Practical Exam Representation

Recipe Site Traffic

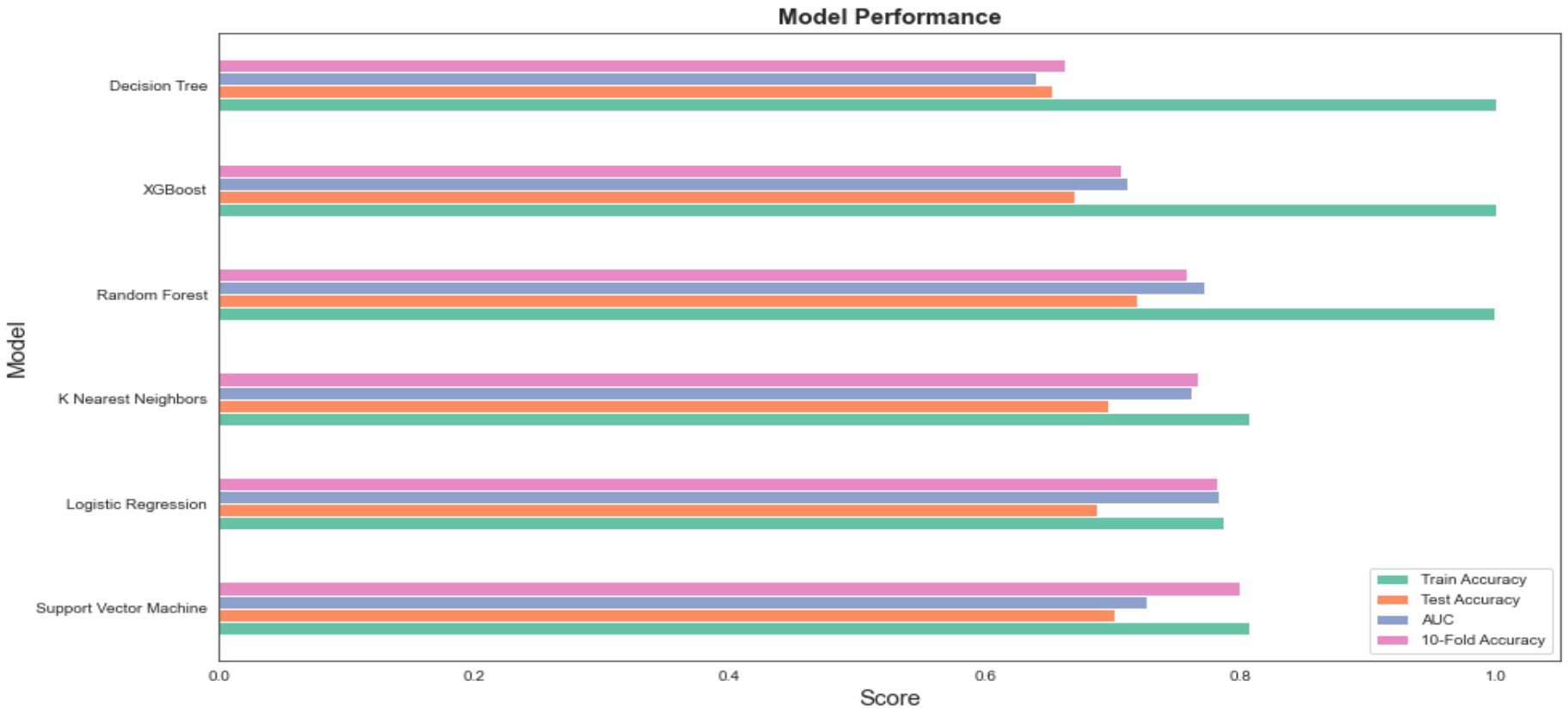
Business Goals

- Tasty Bytes has established for over two years and is a search engine platform for food recipes.
- Creating more attractive recipes will bring more subscribers for the business.
- The product team wants to
  - Determine which recipes will be popular with 80% accuracy.
  - Reduce the probability of showing less popular recipes.

Model Development

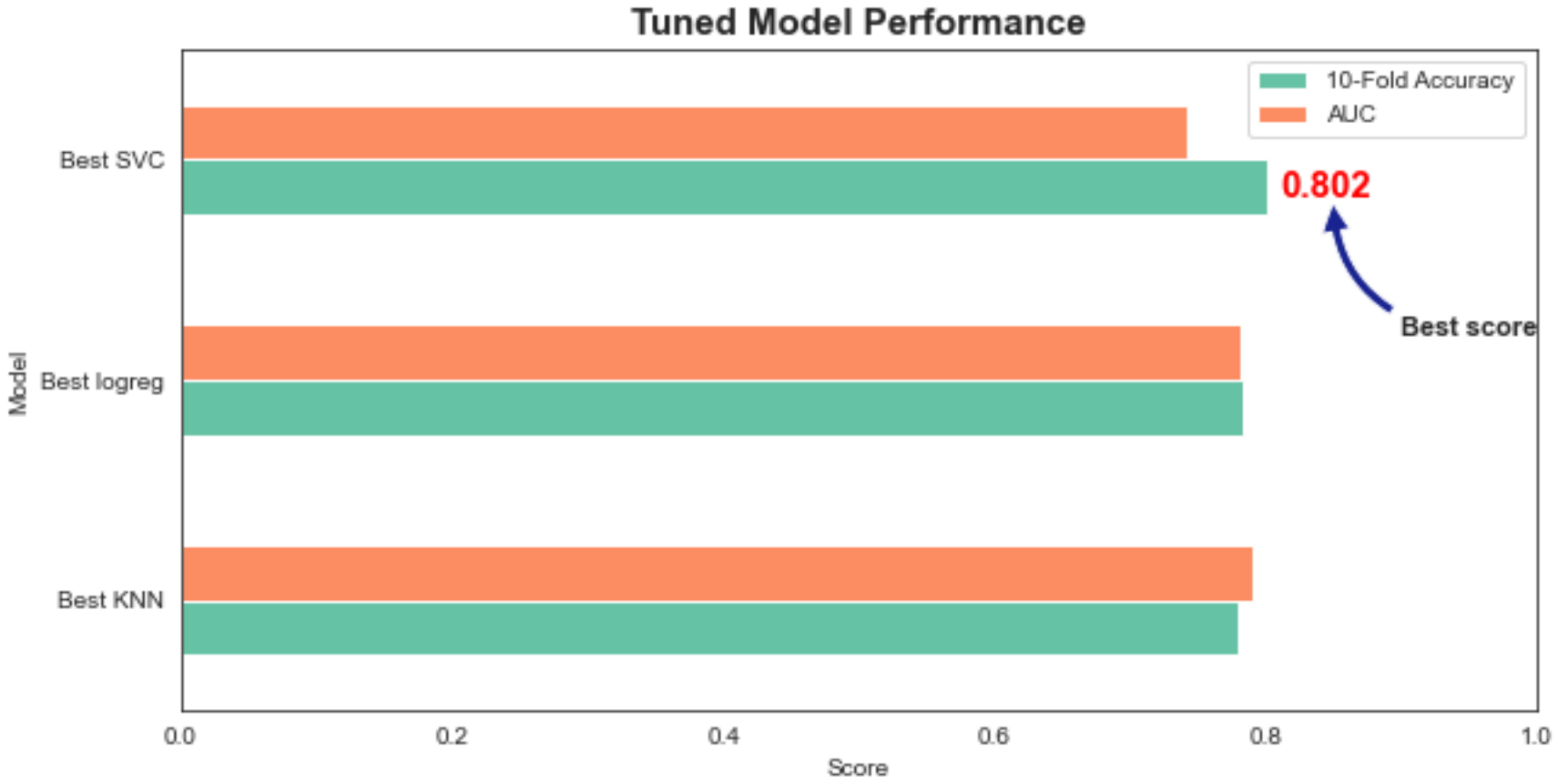


Model Performance



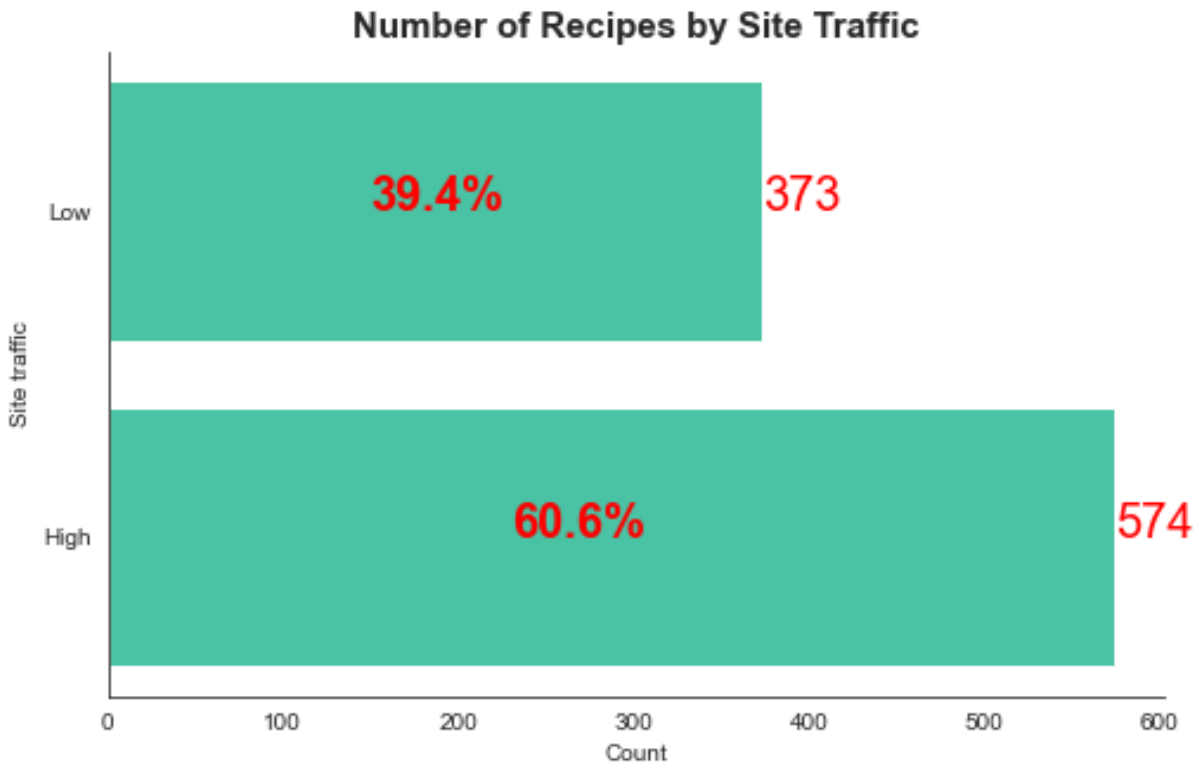
- Based on AUC and Average Accuracy score using 10-fold cross-validation, **Logistic Regression, K Nearest Neighbors,** and **Support Vector Machine** have better performance.
- Though, these models are a little overfitting.

Hyperparameter Tuning



- Use **GridSearchCV** and **RandomizedSearchCV** to tune hyperparameters from each model.
- **Support Vector Machine** provides the highest accuracy score (80%).

AUC vs Accuracy



- **SVC** has higher **Accuracy** but lower **AUC**.
- However, the target variable data is pretty balanced (60% vs 40%), so Accuracy score is a better metric.

✓ Accuracy

Recommendations

- First, collect more data since the original data is kind of small.
- More importantly, collect data more correctly because there are many inappropriate values in the dataset.
- Collect more features such as price, time to make a recipe, and other nutrients like fat.
- Besides building learning models, conduct A/B test to check if some other features have an impact on the site traffic.

Thank you for watching