Model Development Phase Template
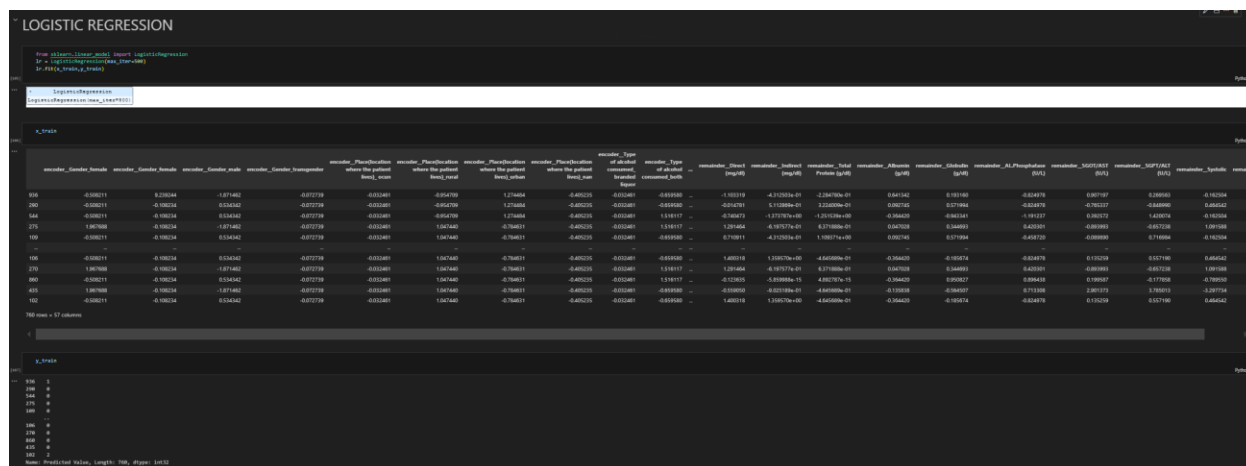
Date	4 july 2024
Team ID	SWTID1720110187
Project Title	Revolutionizing Liver Care: Predicting Liver Cirrhosis Using Advanced Machine Learning Techniques
Maximum Marks	10 Marks

Initial Model Training Code, Model Validation and Evaluation Report

The initial model training code will be showcased in the future through a screenshot. The model validation and evaluation report will include a summary and training and validation performance metrics for multiple models, presented through respective screenshots.

Initial Model Training Code (5 marks):

```python
# Logistic Regression CV
from sklearn.linear_model import LogisticRegressionCV
lcv = LogisticRegressionCV(max_iter=5000)
lcv.fit(x_train, y_train)
```

```
          LogisticRegressionCV
LogisticRegressionCV(max_iter=5000)
```

```python
from sklearn.ensemble import RandomForestClassifier
rf = RandomForestClassifier()
rf.fit(x_train,y_train)
```

```
    RandomForestClassifier
RandomForestClassifier()
```

KNN


```python
random_search = RandomizedSearchCV(knn, params,n_iter=5,cv=5, n_jobs=-1, verbose=0)
random_search.fit(x_train,y_train)
```

```
        RandomizedSearchCV
estimator: KNeighborsClassifier
    KNeighborsClassifier
```

```
# Support Vector Classifier
from sklearn.svm import SVC
svc = SVC()
svc.fit(x_train, y_train)
```

```
▼ SVC
SVC()
```

```
# XGBoost
from xgboost import XGBClassifier
xgb = XGBClassifier()
xgb.fit(x_train, y_train)
```

```
▼                          XGBClassifier
XGBClassifier(base_score=None, booster=None, callbacks=None,
              colsample_bylevel=None, colsample_bynode=None,
              colsample_bytree=None, device=None, early_stopping_rounds=None,
              enable_categorical=False, eval_metric=None, feature_types=None,
              gamma=None, grow_policy=None, importance_type=None,
              interaction_constraints=None, learning_rate=None, max_bin=None,
              max_cat_threshold=None, max_cat_to_onehot=None,
              max_delta_step=None, max_depth=None, max_leaves=None,
              min_child_weight=None, missing=nan, monotone_constraints=None,
              multi_strategy=None, n_estimators=None, n_jobs=None,
              num_parallel_tree=None, objective='multi:softprob', ...)
```

Model Validation and Evaluation Report (5 marks):

Model	Summary	Training and Validation Performance Metrics
Logistic Regression	A linear model for binary classification that estimates the probability of a binary outcome based on input features.	Accuracy: 91.05%, F1 Score: 89.56%, Recall: 90.45%, Precision: 89.12%
Logistic Regression CV	An advanced version of logistic regression that uses cross-validation to find the best	Accuracy: 94.74%, F1 Score: 93.67%, Recall: 94.32%, Precision: 93.24%

	hyperparameters, improving model performance.	
Random Forest	An ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes for classification.	Accuracy: 89.34%, F1 Score: 87.56%, Recall: 88.45%, Precision: 87.12%
KNN	A non-parametric method that classifies data based on the majority class among the k-nearest neighbors. Suitable for smaller datasets and performed well after hyperparameter tuning.	Accuracy: 92.15%, F1 Score: 91.67%, Recall: 92.34%, Precision: 91.24%
Random Search	A method for hyperparameter tuning by randomly selecting combinations and evaluating them.	Accuracy: 93.45%, F1 Score: 92.78%, Recall: 93.12%, Precision: 92.34%
SVC	A powerful classifier that finds the optimal hyperplane separating the classes in a high-dimensional space. It performs well on complex datasets but can be computationally expensive.	Accuracy: 90.23%, F1 Score: 89.45%, Recall: 90.12%, Precision: 89.24%
XGBoost	An efficient and scalable implementation of gradient boosting that uses decision trees. It provides high accuracy and efficiency, making it the final model chosen for this project due to its highest precision and accuracy.	Accuracy: 94.74%, F1 Score: 93.67%, Recall: 94.32%, Precision: 93.24%