Table A1: Comparison of Machine Learning Models

Models	**Advantages**	**Disadvantages**
k-nearest neighbor (k-NN)	[left=0pt, nosep]Nonparametric Zero cost in learning process Intuitive approach Robust to outliers	[left=0pt, nosep]Expensive computation for large datasets Hard to interpret results Performance depends on number of dimensions Lack of explicit model training
Support vector machine (SVM)	[left=0pt, nosep]Utilizes predictive power of linear combinations Good prediction in various situations Low generalization error	[left=0pt, nosep]Weak handling of mixed data types Sensitive to tuning parameters and kernel choice Slow training for large datasets
Decision Trees	[left=0pt, nosep]Tolerance to correlated inputs Highly interpretable (single tree) Handles missing values Works with numerical and categorical data	[left=0pt, nosep]Cannot work on combinations of features Relatively less predictive in many situations Prone to overfitting
Logistic regression	[left=0pt, nosep]Provides logistic probability model Easy to interpret Provides confidence interval	[left=0pt, nosep]Doesn't handle missing continuous variable values Suffers multicollinearity Sensitive to extreme continuous variable values
Naïve Bayes	[left=0pt, nosep]Suitable for small training sets Easily obtain probability for prediction Simple and straightforward to use	[left=0pt, nosep]Prone to bias with more training data Assumes all features are independent and equally important Sensitive to data preparation
Neural networks	[left=0pt, nosep]Generally good prediction Tolerance to correlated inputs Incorporates predictive power of different input combinations	[left=0pt, nosep]Not robust to outliers Susceptible to irrelevant features Difficult with big data and complex models

Table A2: Hyperparameter Grid Definition

Model	**Hyperparameter Grid**
KNN	'n_neighbors': [3,5,7], 'weights': ['uniform', 'distance'], 'metric': ['euclidean', 'manhattan', 'minkowski']
MLP	'hidden_layer_sizes': [(50,), (100,), (150,)], 'activation': ['relu', 'tanh'], 'solver':['adam'], 'learning_rate':['constant'], 'power_t':[0.5], 'alpha':[0.0001], 'max_iter':[10000], 'early_stopping':[False], 'warm_start':[False]
RF	'n_estimators': [10, 50, 100], 'max_depth': [5, 10]
CART	'max_depth': [5, 10], 'min_samples_split':[2,3]
NB	‘var_smoothing' : [1e-11, 1e-10, 1e-9]
LDA	‘solver': ["svd","lsqr"]
QDA	‘reg_param':[0.1,0.2,0.3,0.4,0.5]
LR	'C': [0.001, 0.01, 0.1, 1.0]
Aboost	'n_estimators': [50, 100, 200], 'learning_rate': [0.01, 0.1, 1.0]