Machine Learning Process

Gather Data			API, DB, csv, etc.
Clean Data			Fill NaN
			Convert categorical to numerical
			Scaling
	Train/Test Data		Cross-Validation (ways to split)
Feature Selection	Filter Methods		Too high variance
			Too low variance
			By feature similarity (based on r^2 score)
			Chi-squared test
			Anova
			Correlation coefficient scores
	Wrapper Methods		Forward selection
			Backward elimination
			Stepwise selection
Model Selection	Supervised Learning	Regression	Multiple Linear Regression
			Polynomial Regression
			Decision Tree
			Random Forest
			Support Vector Regression
		Classification	KNN
			Decision Tree
			Random Forest
			Logistic Regression
			SVM
			Kernel SVM
			Naive Bayes
	Unsupervised	Clustering	K-means
			Hierarchical
			Agglomerative
			DBSCAN
		Dimensionality Reduction	PCA
			LDA
			MDS
Model Evaluation	Regression		Mean Squared Error (MSE)
			Root-Mean-Squared-Error (RMSE)
			Mean-Absolute-Error (MAE)
			R^2 or Coefficient of Determination
			Adjusted R^2
			Durbin Watson (~2)
	Classification		Accuracy
			Precision/Recall
			F1-score
			ROC curve/AUC score
			Log-Loss
			Fisher scoring iterations (<10) Logistic Regression
			McFadden values (0.2-0.4) Logistic Regression
			Kappa Values (>.9) Naive Bayes
	Clustering		Domain Knowledge
			Radar Plots
	Dimensionality Reduction		?
Model Improvement	Hyperparameter Tuning		GridSearch
	Regularization (Simplify Model)		Ridge Regression (L2 Regularization)
			Lasso Regression (L1 Regularization)
	Ensemble Methods (Increase complexity)		Bagging
			Boosting
			Stacking
Model Deployment			Deploy