

All code and write ups available on GitHub: [kschuerger/kschuerger-projects \(github.com\)](https://github.com/kschuerger/kschuerger-projects)

## **CASE STUDIES (Python)**

### **Case study 1**

- Task: Predict new superconductors, identify important features
- Key methods: Linear regression
- Additional methods: Regularization techniques – L1/Lasso, L2/Ridge

### **Case study 2**

- Task: Predict patient readmission within 30 days
- Key methods: Logistic regression
- Additional methods: Feature engineering (specifically imputation of missing values)

### **Case study 3**

- Task: Antispam email filter
- Key methods: KNN (K-Nearest neighbors) classifier, Random Forest classifier
- Additional methods: TFIDF (Term frequency, inverse document frequency), NLP (count vectorizer), multinomial Naïve Bayes, multilayer perceptron (MLP)

### **Case study 4**

- Task: Predict bankruptcy
- Key methods: Logistic regression classification with L2 regularization, Random Forest classification, XGBoost
- Additional methods: Randomized cv search
- Evaluation methods: ROC curve (Receiver Operating Characteristic), confusion matrix, F1-score, stratified cross-validation

### **Case study 5**

- Task: Accept or reject firewall request
- Key methods: SVM, SGD
- Additional methods: Kernel tuning (linear, poly, RBF), regularization parameter tuning, Vowpal Wabbit (explored, not used in case study)

### **Case study 6**

- Task: Binary classification model to detect new particle
- Key methods: neural network (TensorFlow, Keras API) – sequential model used
- Additional methods: binary cross entropy, ROC, confusion matrix

## **PUBG MODELS (Python)**

### **PUBG Models 01 EDA**

- Task: Initial EDA of dataset
- Methods: Linear discriminant analysis (LDA), data visualizations, features engineering (impute missing values, scale data, evaluate outliers, create new features)

### **PUBG Models 02 Tasks**

- Task: Initial setup of classification and regression tasks (targets: discrete and continuous, respectively)
- Methods: LDA, random search cv, cross validation, train/test split, PCA, grid search, random forest, linear regression, KNN, decision tree, regularization (Lasso), pipeline, SVC (support vector machine), reduce memory usage

### **PUBG 03 Final**

- Task: Final analysis and evaluation
- Methods: Combination of Models 01 and Models 02 projects.

## **TIME SERIES (R)**

### **Time series Walmart stock**

- Task: Fit model using time series methods, evaluate short-term and long-term forecasts
- Methods: Model ID using AIC, ACF, PACF, univariate and multivariate time series analysis, Ljung-Box test, rolling window RMSE/ASE, ARMA, ARIMA (differencing), vector autoregressive (VAR) model, cross-correlation, multilayer perceptron (MLP, neural network)