

# **Data Visualization Final Project**

The main goal of this project is to apply the Machine Learning methods we studied this semester and reflect the knowledge we have gained since the previous Data Science course.

## **Jupyter Notebooks:**

### **1. 1SB ML Hotel Bookings:**

We were asked to improve our classification notebook from the Introduction to Data Science project, using new algorithms and theories we acquired.

#### **Models and methods used:**

- Random Forest
- KNN
- Logistic Regression
- XGBoost
- Fine-Tuning with Grid Search
- Voting Classifiers (hard and soft voting)
- Stacking

Result: F1 score improved by 0.7%.

### **2. 1SB ML Fashion-MNIST:**

A continuation to the famous MNIST Digits dataset, the Fashion-MNIST is a multiclass dataset that contains images of clothing items.

#### **Models and methods used:**

- Image Pre-Processing (transformations: grayscale, black and white)
- Logistic Regression (softmax, OvR, OvO)
- KNN
- Decision Tree
- Random Forest
- PCA
- Extra Trees
- AdaBoost
- XGBoost
- Fine-Tuning with Grid Search
- Voting Classifiers (hard and soft voting)
- Data Augmentation

Result: 89.57% Accuracy.

### 3. **1SB ML Dogs vs. Cats:**

This image dataset is made up of pictures of dogs and cats. This classification task was challenging to tackle with Machine Learning algorithms.

#### **Models and methods used:**

- Image Pre-Processing (progressive resizing, transformations: color, grayscale, canny)
- Logistic Regression
- KNN
- Decision Tree
- Random Forest
- K-Means Clustering (for dimensionality reduction)
- PCA
- Extra Trees
- AdaBoost
- XGBoost
- Fine-Tuning with Grid Search
- Voting Classifiers (hard and soft voting)
- Stacking

Result: 67.24% Accuracy.

### 4. **1SB Interpersonal Physical Alignment:**

This notebook uses real data from an academic study. The data contains recorded hand motions that were gathered using the Leap Motion controller. Each participant was recorded in 3 different states: alone, spontaneous and sync. The goal is to train a Machine Learning algorithm to correctly predict the state when given unseen data.

#### **Models and methods used:**

- Data Preparation
- Logistic Regression (softmax)
- KNN
- Decision Tree
- Random Forest
- PCA
- AdaBoost
- XGBoost
- Fine-Tuning
- Voting Classifiers (hard and soft voting)

Result: 86.74% Accuracy.