Work summary

By: Sahar Tuvyahu

The methods used in this work and their explanations:

- **Models:**
  'Weak' learners:
  - KNN
  - Logistic Regression
  - Decision Tree

  Ensemble learning:
  - Random Forest: estimate the result of a collection of decision trees.
  - Voting Classification:
  Hard: aggregate the predictions of each classifier and predict the class that gets the most votes.
  Soft: predict the class with the highest-class probability, averaged over all the individual classifiers.
  - Stacking: train a model to perform the aggregation of voting.

  Boosting:
  - Adaboosting: each successor model fits according to the weight of its predecessor's mistakes.
  - Gradient Boosting: same as adaboost, but this method tries to fit the new predictor to the residual errors made by the previous predictor.
  - XGBoosting: Extreme Gradient Boosting

  Clustering:
  - K-means: classification using k clusters as 'starting point' on what groups of data the model should find

- **Dimensionality Reduction:**
  - PCA: choosing a lower dimensional hyperplane of the data that preserve X value of the original data's variance. (Number of features that preserve that variance)

Classification Enhancements:

- Random Forest, Voting, Stacking
- Adaboost, XGBoost

FMNIST:

- KNN, Logistic Regression, Decision Tree
- Random Forest, Voting, Stacking
- Adaboost, XGBoost
- PCA

Cats vs Dogs:

- KNN, Logistic Regression, Decision Tree
- Random Forest, Voting, Stacking
- Adaboost, XGBoost
- K-means
- PCA

Hands:

- KNN, Logistic Regression, Decision Tree
- Random Forest, Voting, Stacking
- Adaboost, XGBoost
- K-means
- PCA