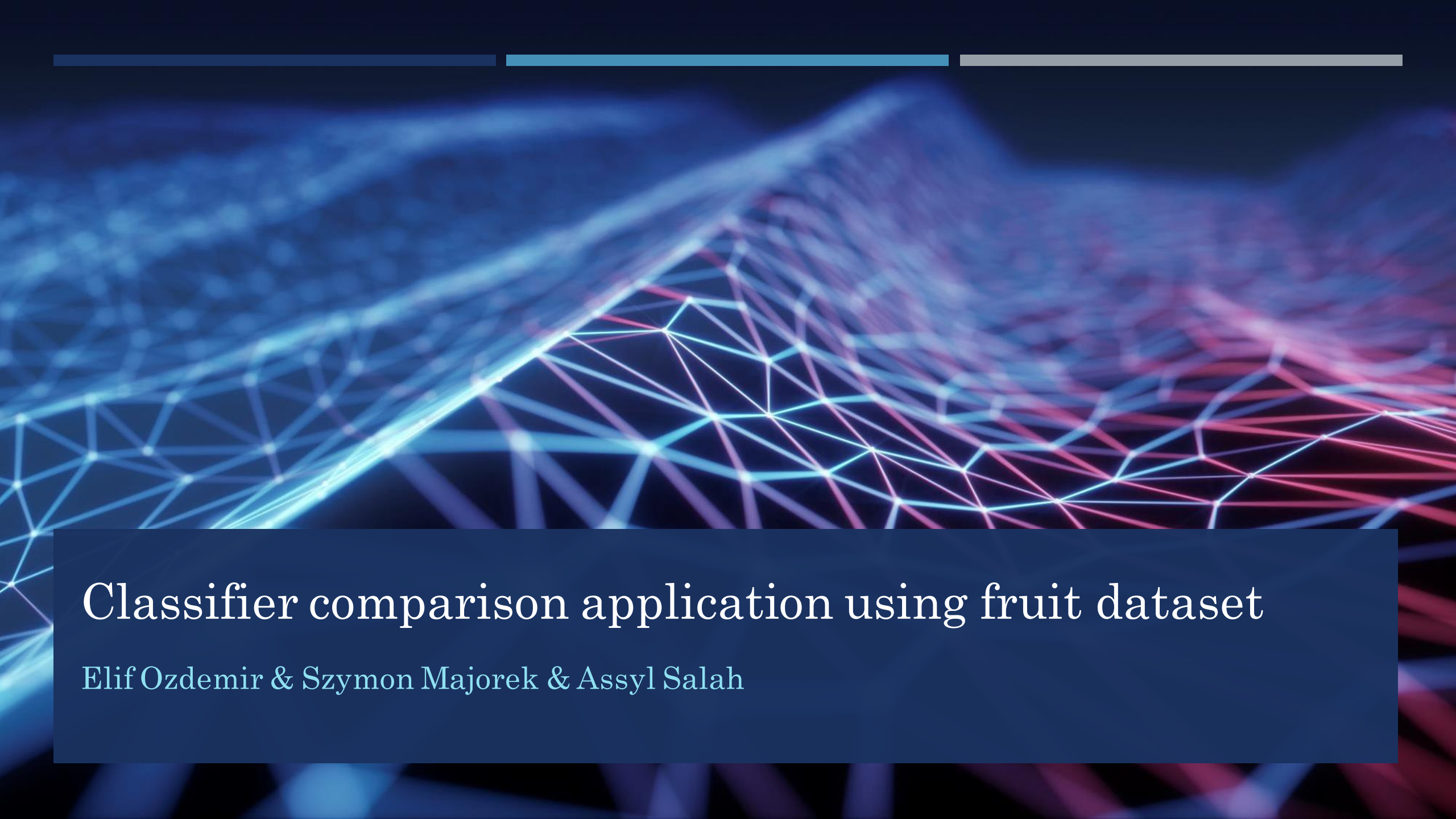
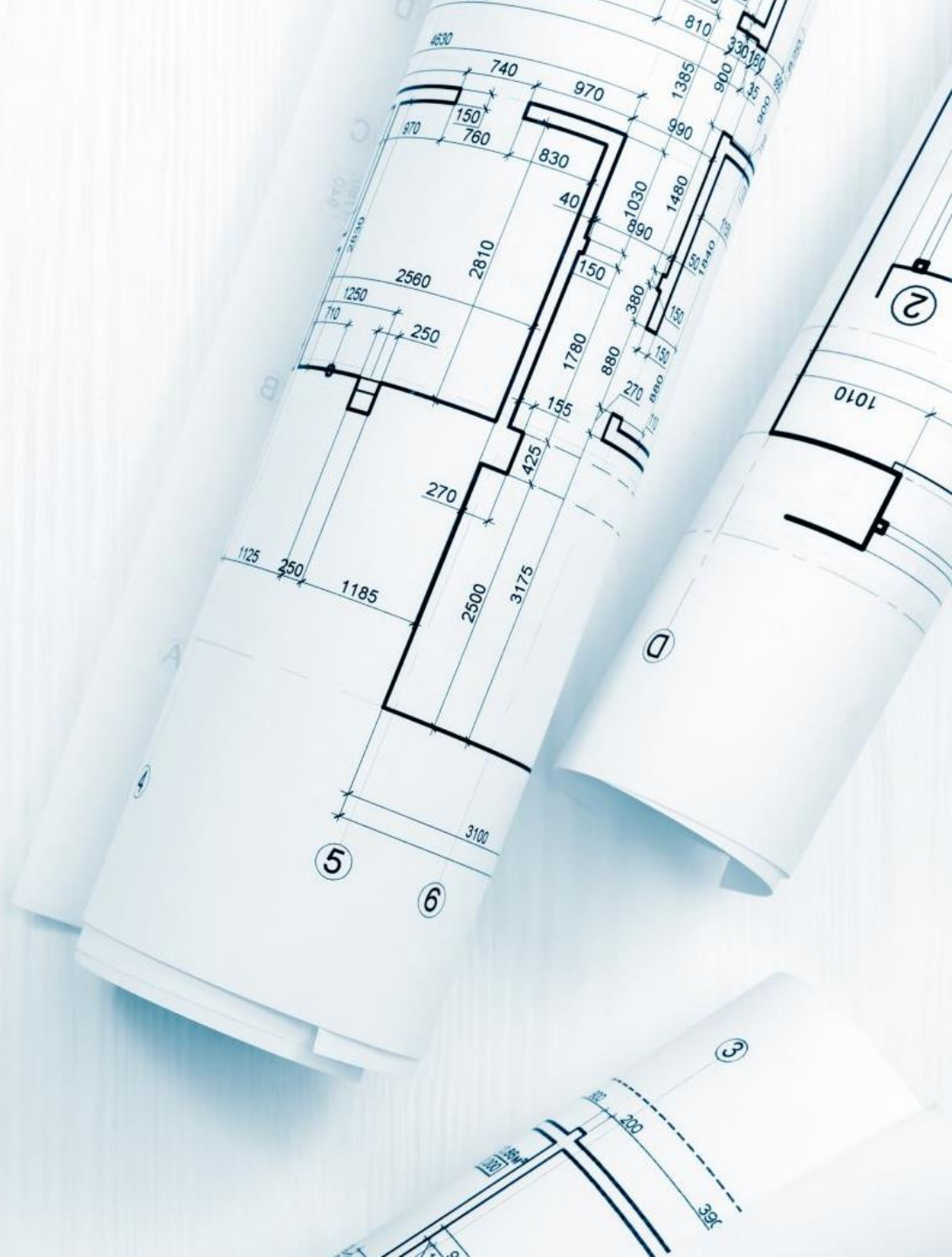


Classifier comparison application using fruit dataset

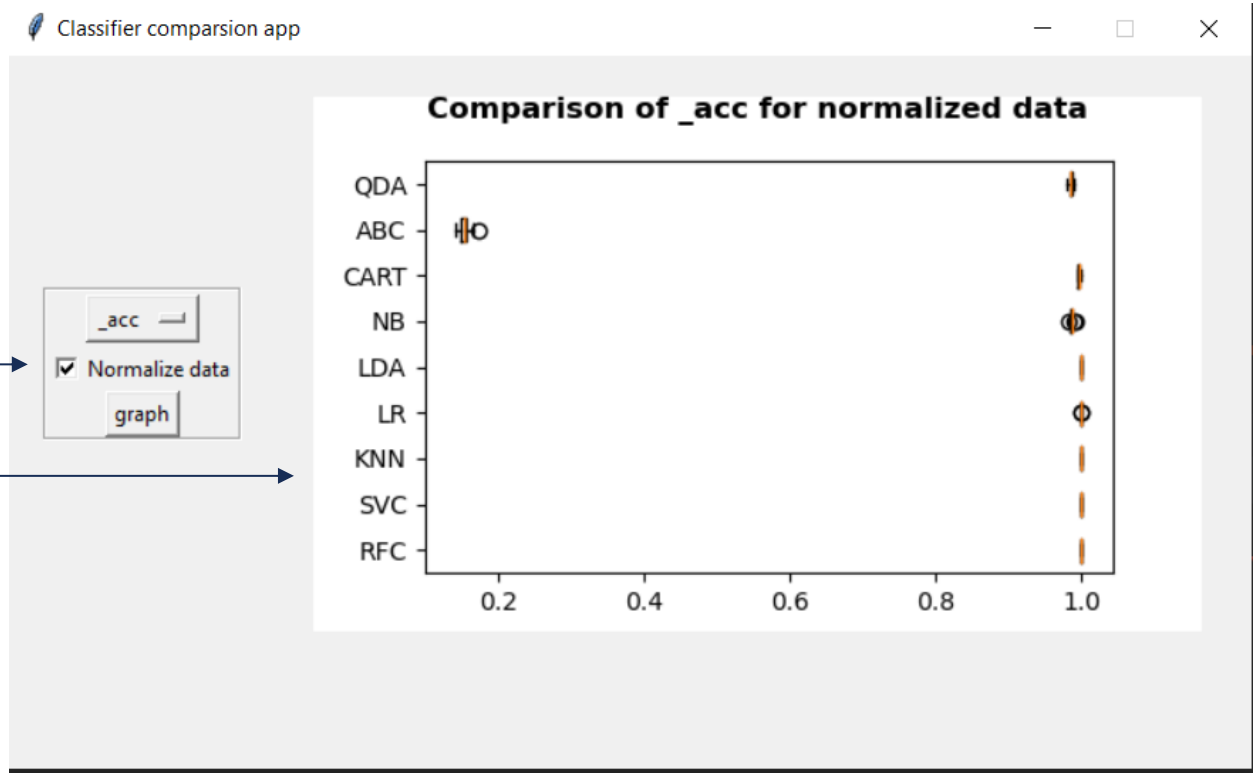
Elif Ozdemir & Szymon Majorek & Assyl Salah

- Application Description
- Application Purpose
- Dataset Choice
- Chosen Models
- Application User Interface
- Graphs
- Summary

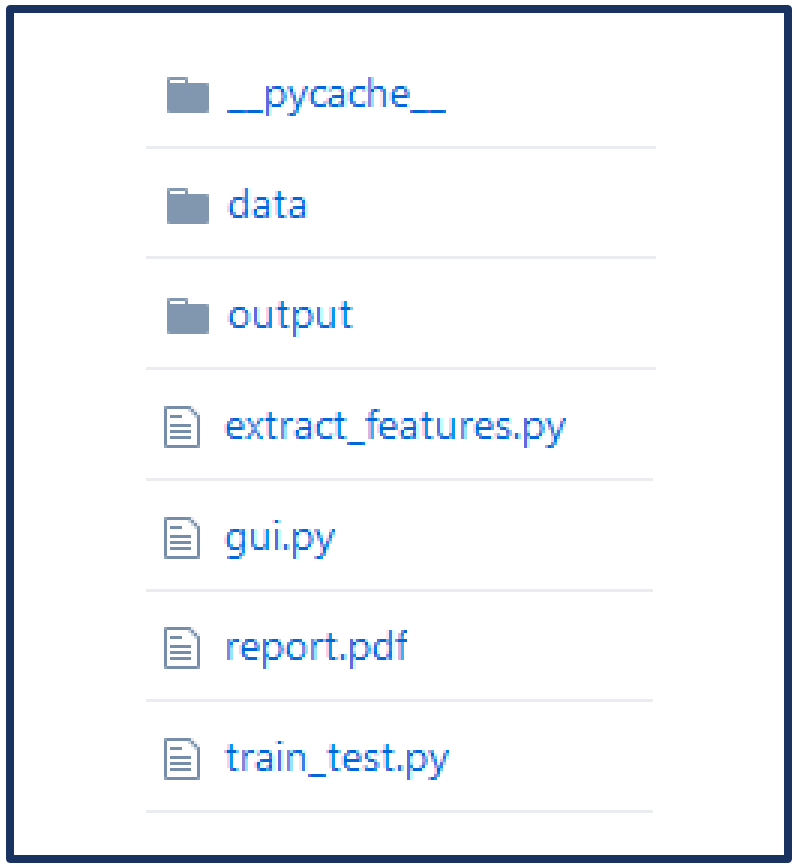


APPLICATION DESCRIPTION

- We've aimed to keep the application as simple as possible, hence the minimalistic interface.
- It has 2 main areas.
  - Control panel
  - Graph display area
- In the control panel we can set metric for which we want to compare classifiers, as well as whether we want to use normalized data.



APPLICATION – BEHIND THE SCENES



The application is composed of few simple python scripts responsible for different tasks.

- *gui.py* is only responsible for the layout and showing selected information to the user in the form of boxplot.
- *extract_features.py* takes fruit images (located in *data* directory) as an input and produces HDF5 files (*output* directory).
- *train_test.py* Takes beforementioned HDF5 files as an input and trains, as well as tests given machine learning models.

The natural flow of logics is extract_features -> train_test -> gui

APPLICATION PURPOSE

- Image recognition, in the context of machine vision, is the ability of software to identify objects, places, people, writing and actions in images. Computers can use machine vision technologies in combination with a camera and artificial intelligence software to achieve image recognition.
- The main purpose of this project application is to test efficiency of different Python machine learning models based on different criteria.
- Our application will give an idea to the user about the accuracy, error rate, fitting and scoring time which define the success rate of the models.

Criteria with which the user will be able to compare different models against each other

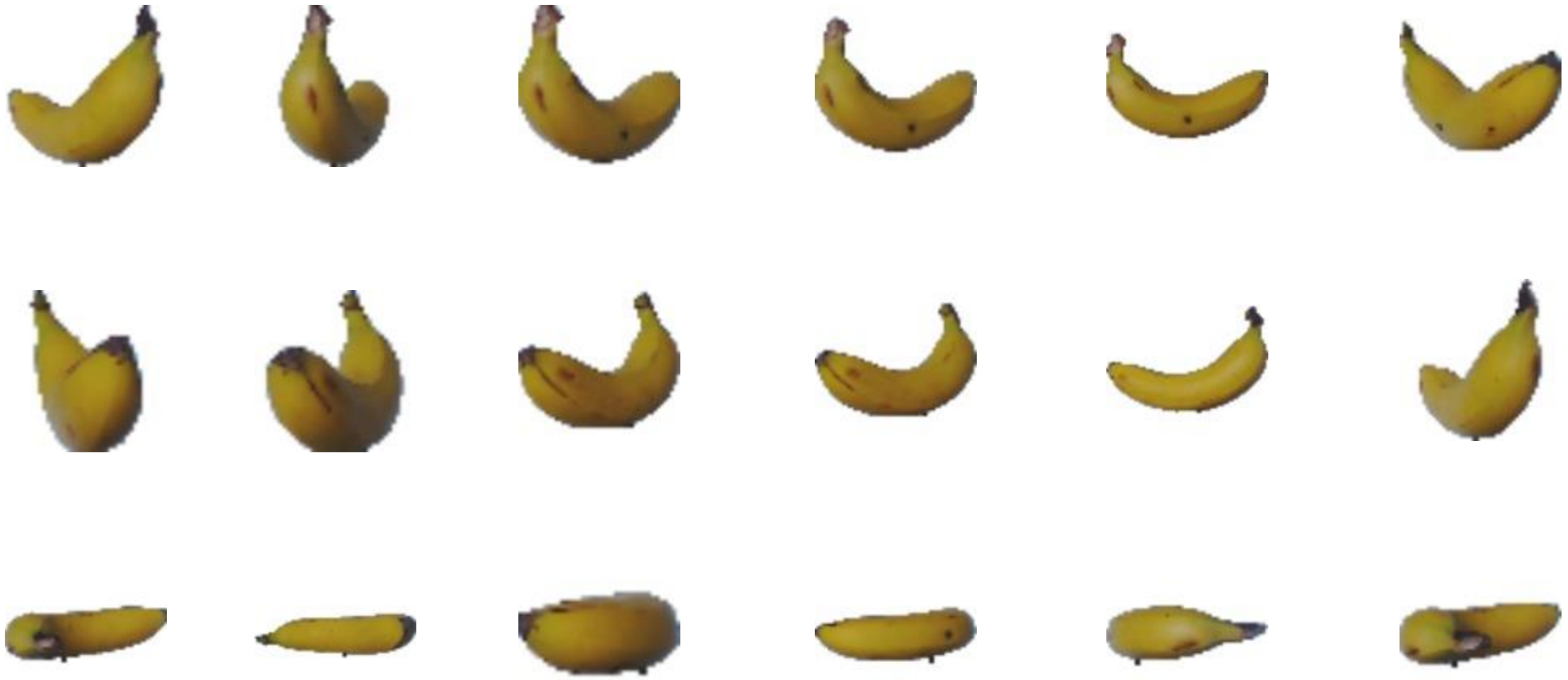
- Accuracy
- Mean squared error
- Fit time
- Score time

DATASET CHOICE

DIVERSITY OF FRUITS/ VEGETABLES

- 28,666 images in total
- 44 species
- 21,444 -training set size
- 7,222 -testing set size
- More than 490 pictures per a fruit in training set
- More than 150 pictures per a fruit in validation set
- Good lighting
- Stable background
- Fixed object size
- 100x100 pixels

CAMERA ANGLE DIVERSITY (360 DEGREES)



CHOSEN MODELS

- Logistic Regression
- Linear Discriminant Analysis
- K-Nearest Neighbors
- Decision Tree
- Random Forests
- Gaussian Naive Bayes
- Support Vector Machine
- Quadratic Discriminant Analysis
- Ada Boost classifier

CROSS VALIDATION SCORE BASED ON

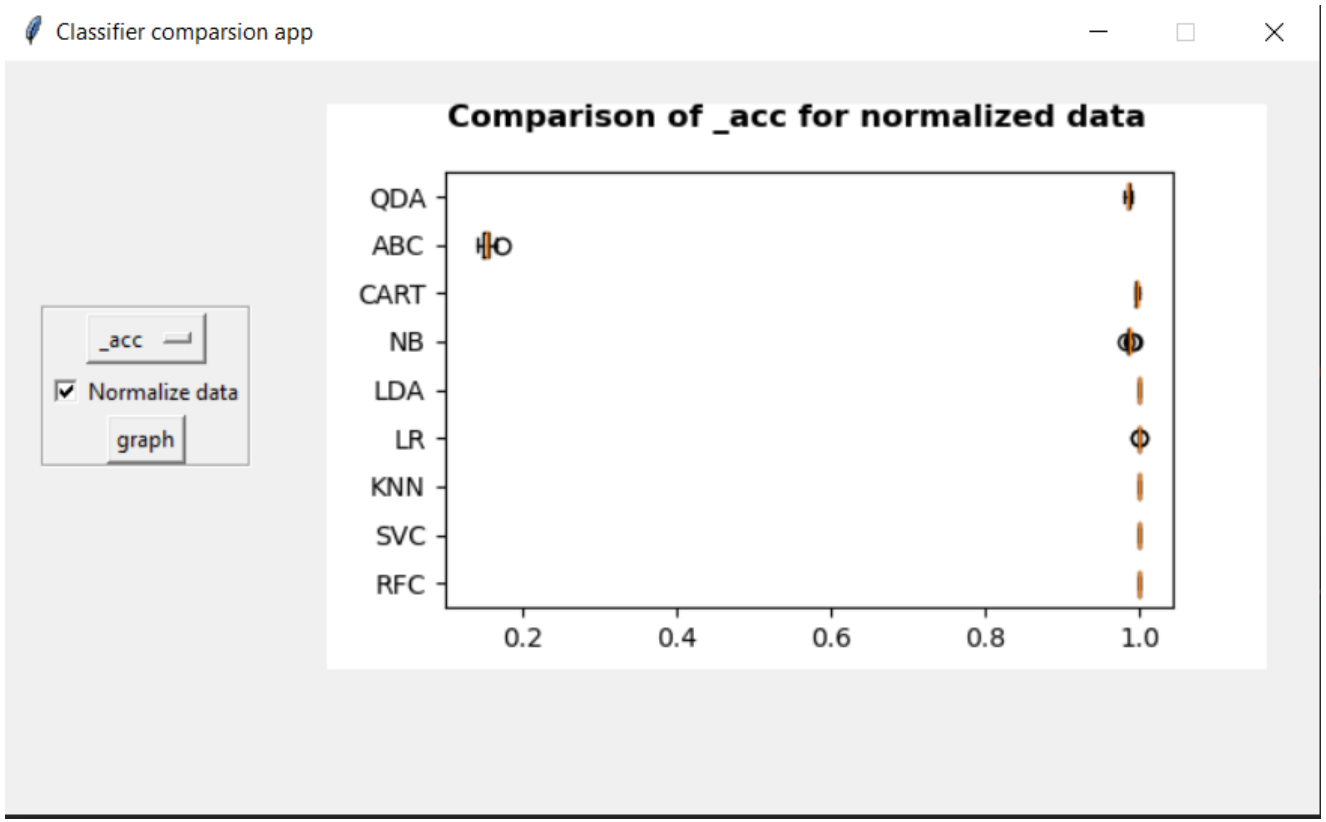
Accuracy

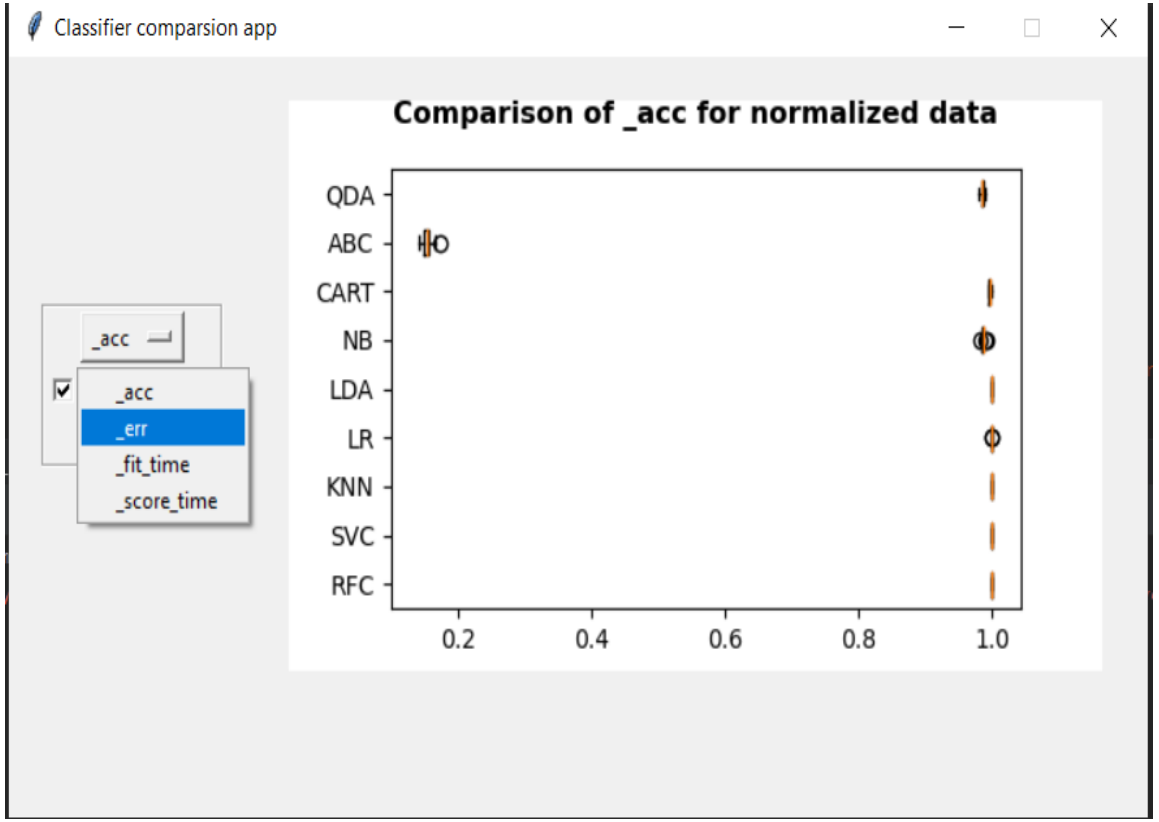
Mean squared error

Fit time

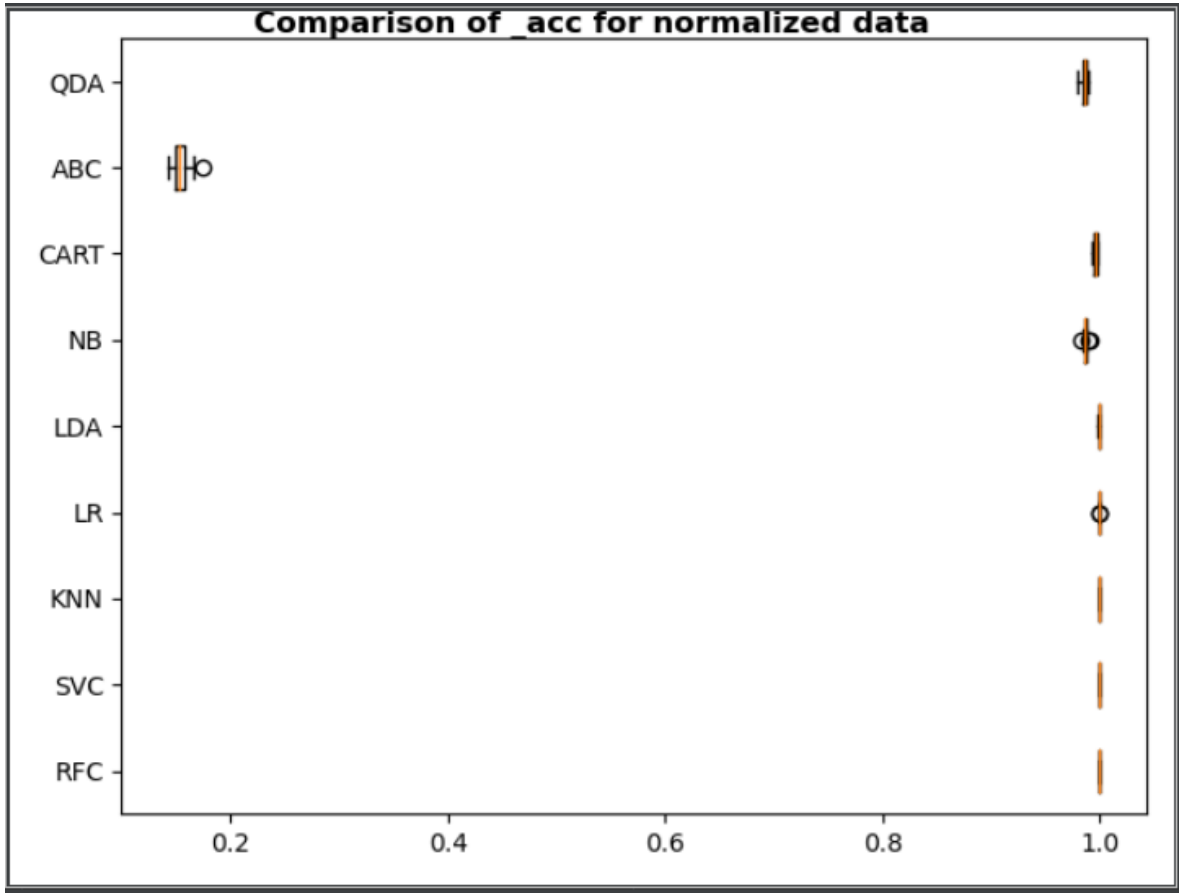
Score time

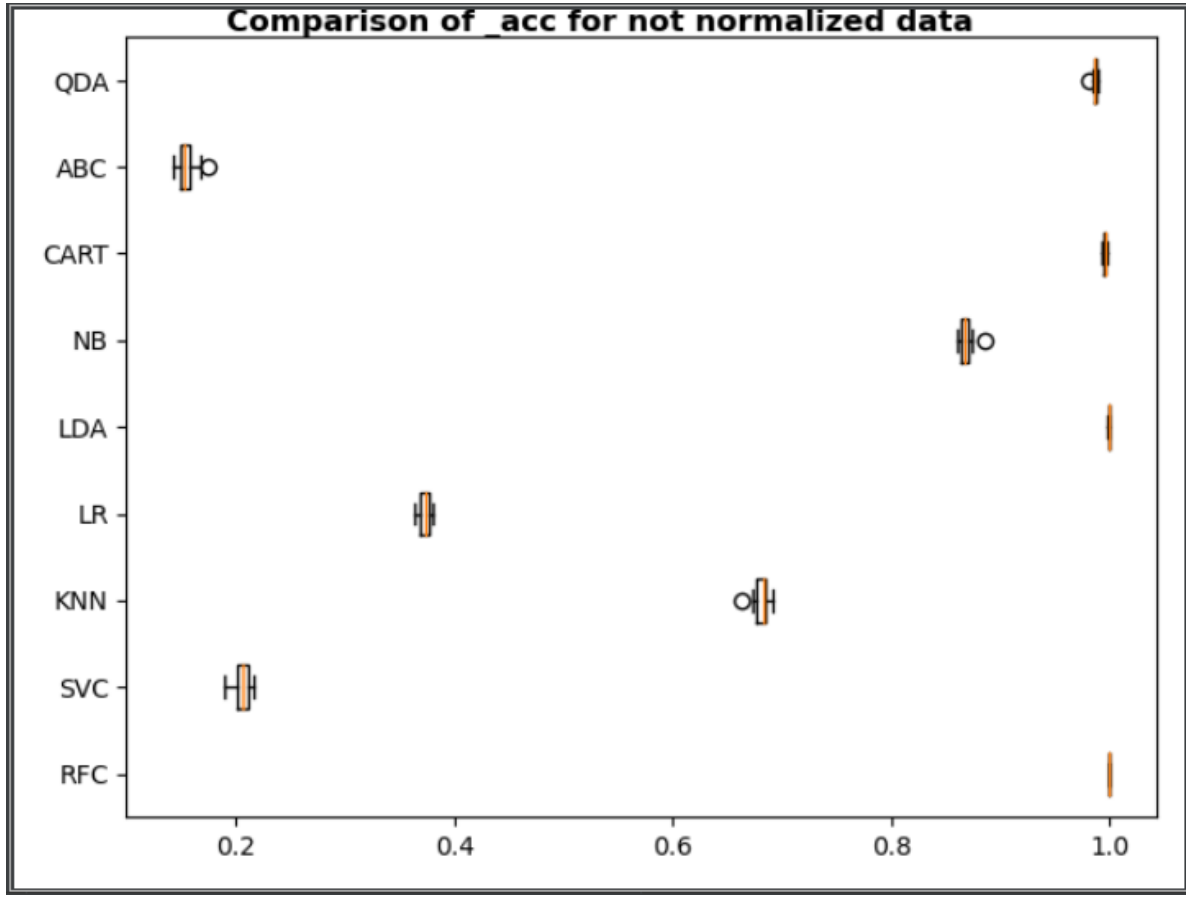
GRAPHICAL USER INTERFACE





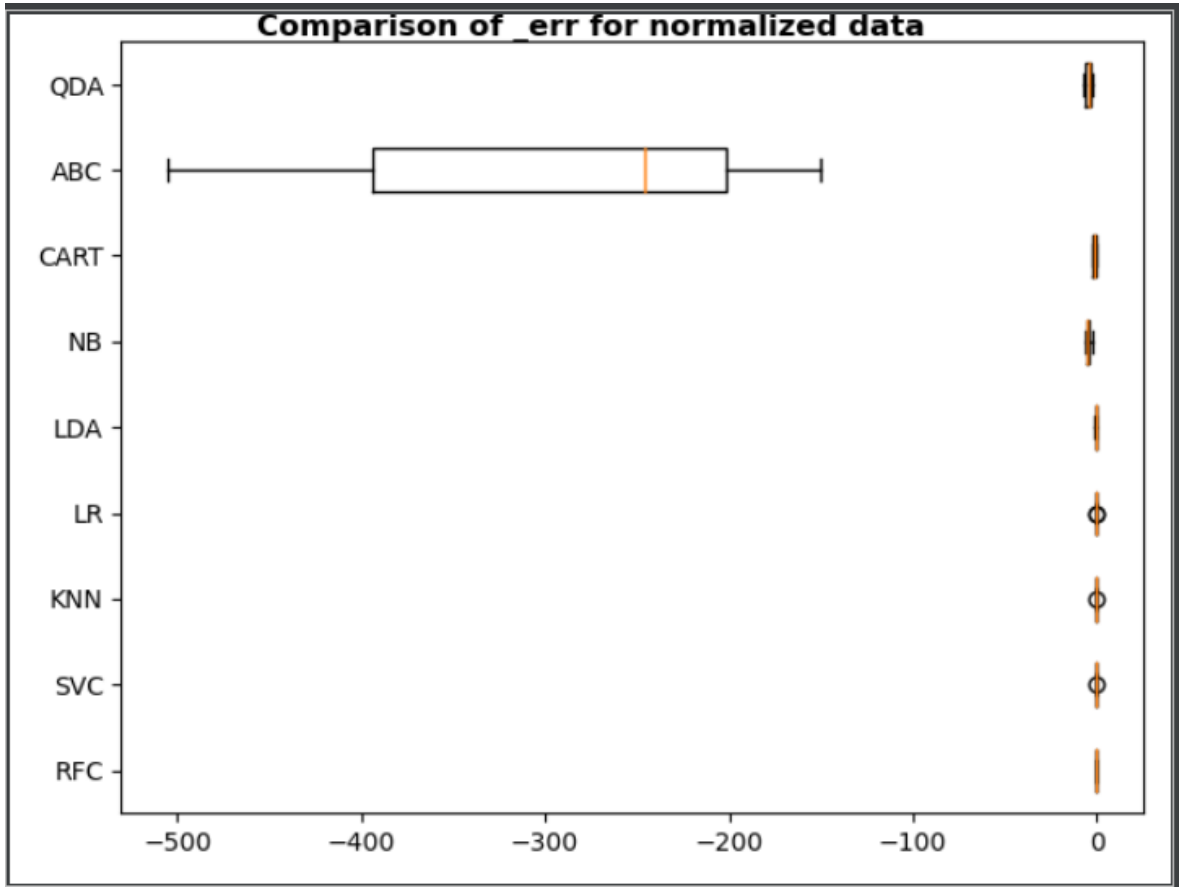
GRAPHS

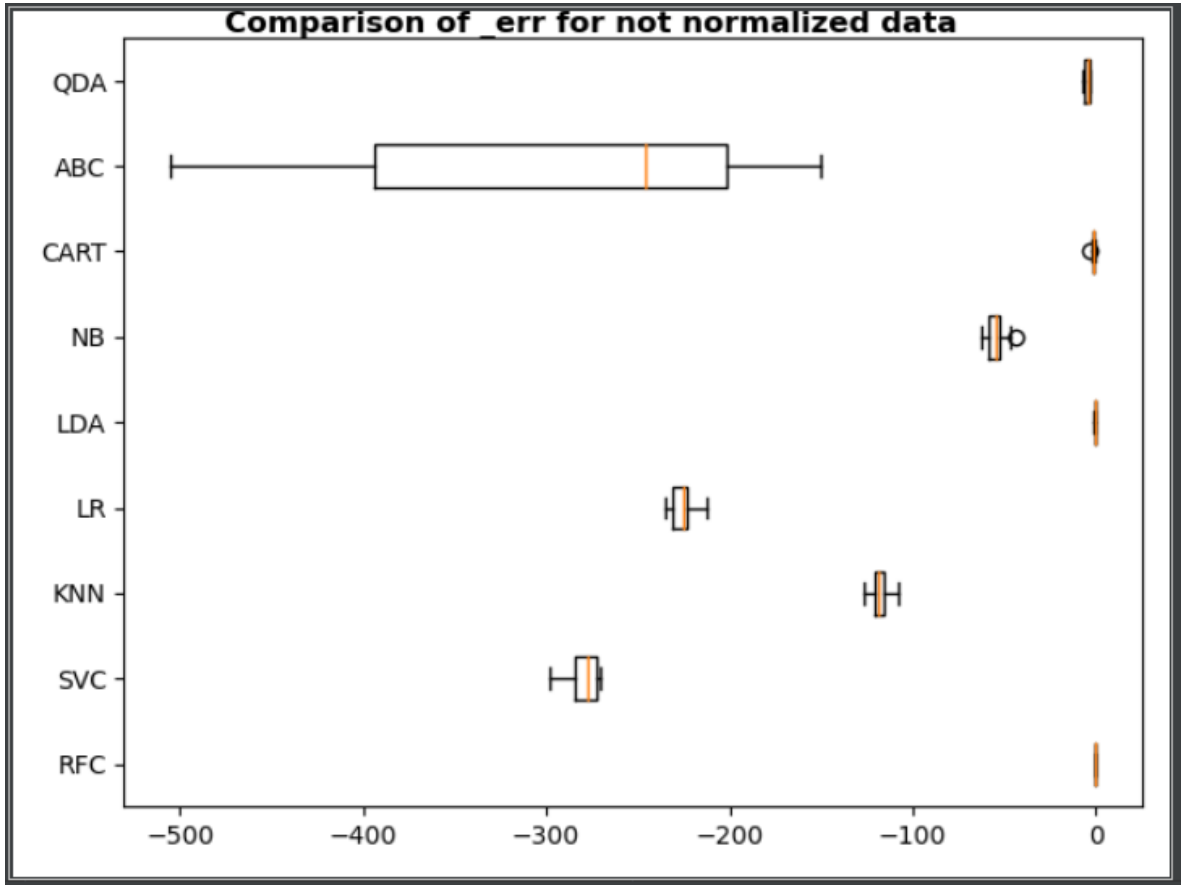




Ada Boost classifier gives low accuracy
But normalization of the data helps in favor of the other classifiers. .

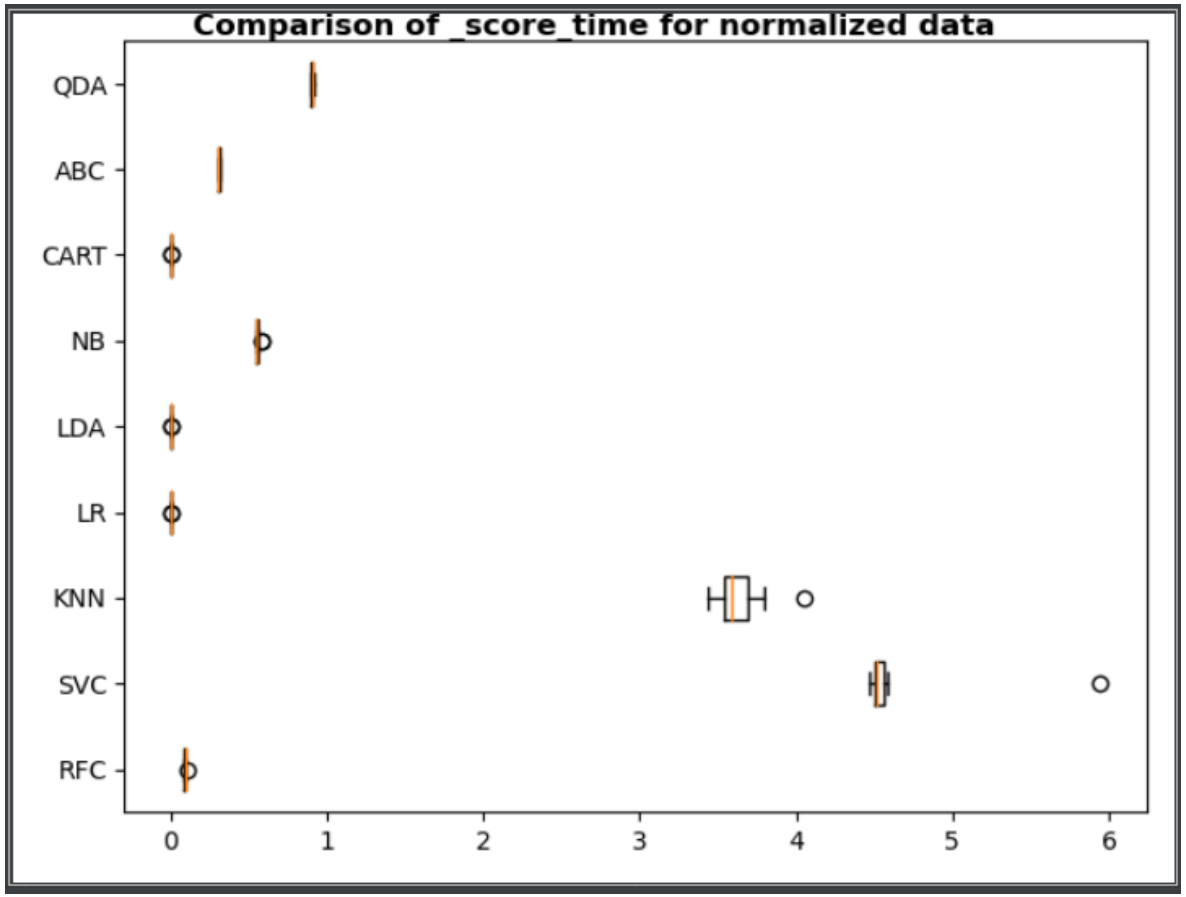
Random Forests, Linear Discriminant Analysis and Decision Tree are highly successful. However, ABC and SVC have very low accuracy.

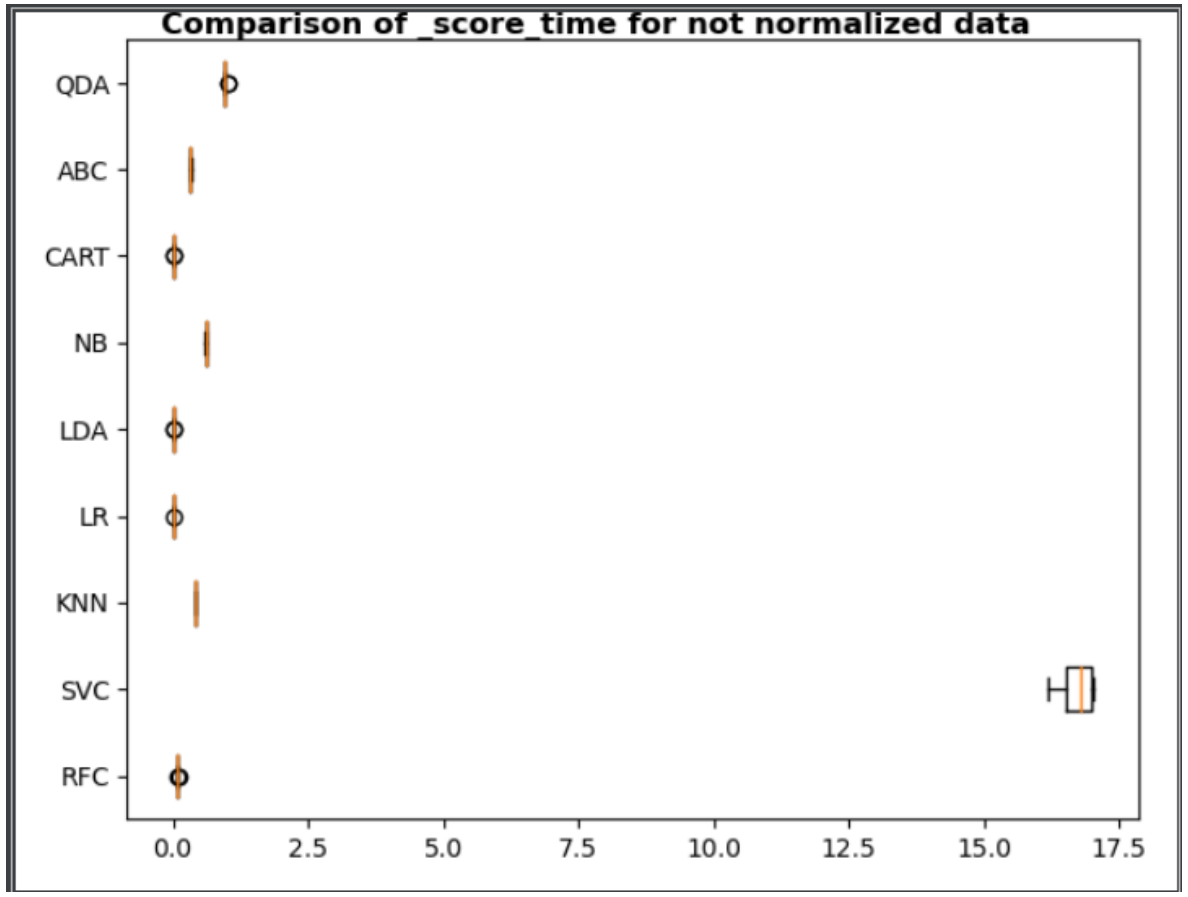




The amount of errors is close to 0 in the majority of classifiers. However, Ada Boost is an exception

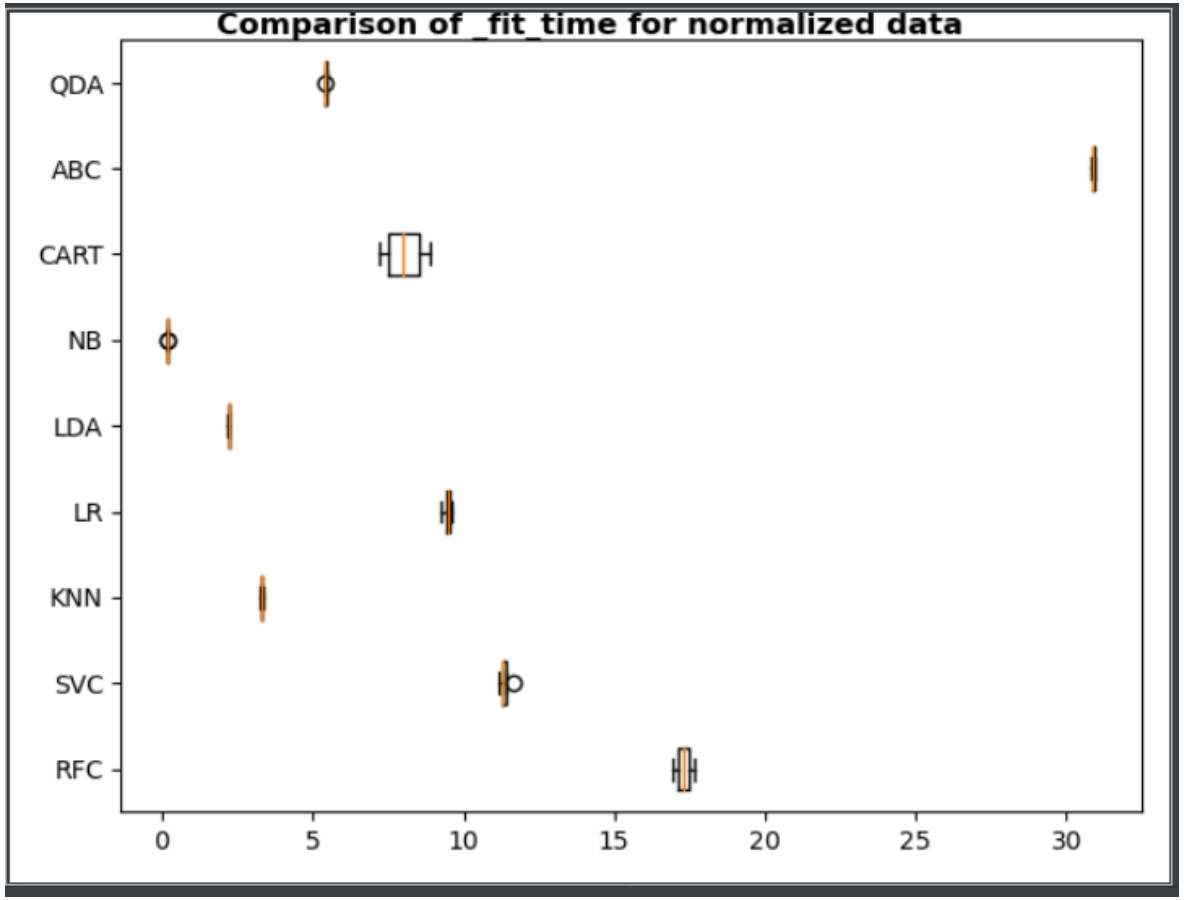
For Quadratic Discriminant Analysis, Decision Tree, Linear Discriminant Analysis and Random Forests classifiers error rate is low

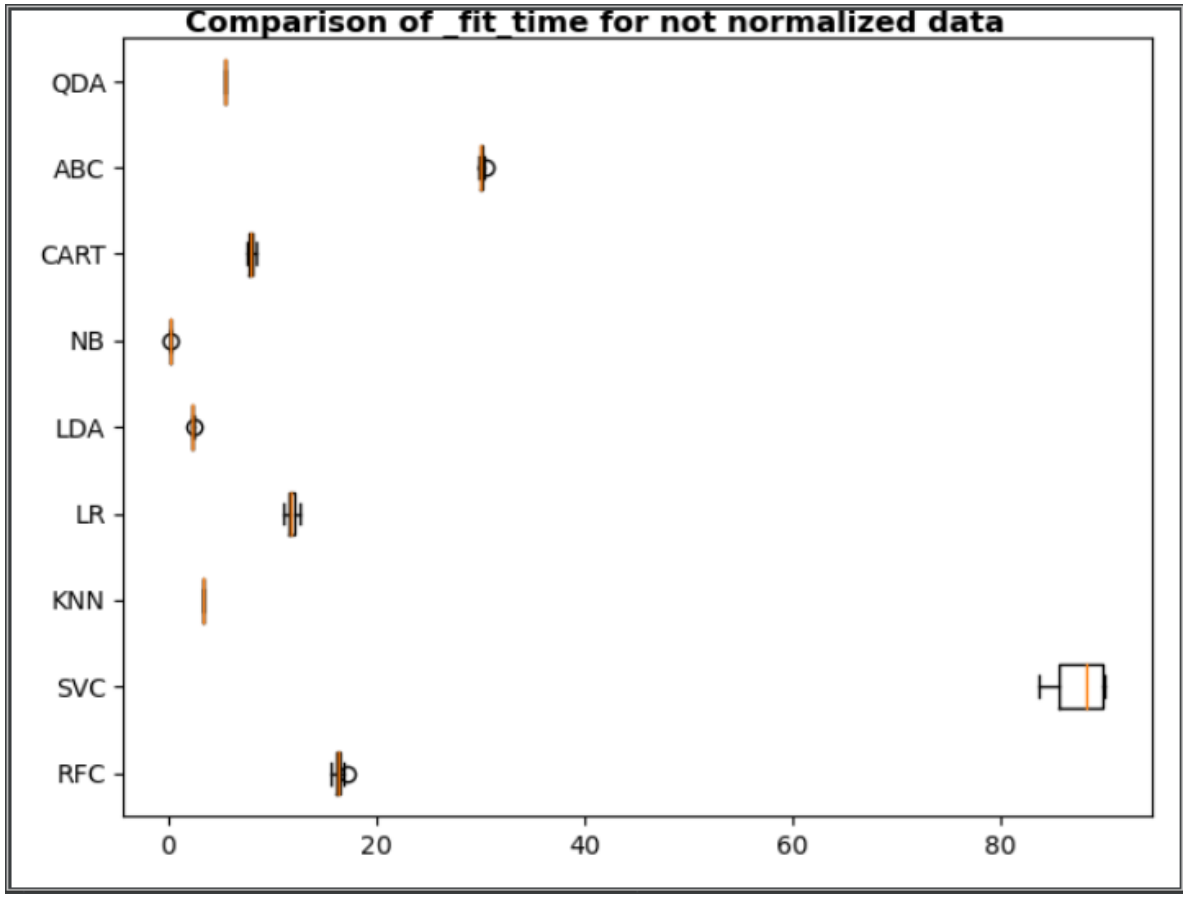




Support Vector Machine Model has the highest scoring time
The second worst result is for K-Nearest Neighbors Model.

Scoring time is the worst for Support Vector Machine Model.





Gaussian Naive Bayes is the fastest in fitting the data.
Ada Boost classifier is the slowest.

Support Vector Machine Model takes at least 4 times more time compared to other classifiers. Except Support Vector Machine and Ada Boost classifier all the models fit the data at 20 seconds

SUMMARY

MAIN TAKEAWAYS

The app finds out accuracy, error rate, fitting and scoring time depending on:

- quality of the data
- quantity
- normalization

Best in non-normalized data

- Linear Discriminant Analysis
- Quadratic Discriminant Analysis
- Decision Tree
- Random Forests

Worst in the normalized data

- Ada Boost classifier

A QUESTION

Which of the fruits was upside down?

THANK YOU