# Forest cover type

The task involved building and testing different models on a given dataset to predict forest cover type based on cartographic variables.

The entire task was divided into several files. Jupyter Notebook was used to build the models, but the file added to the GitHub platform has a .py format. The first file, named "Heuristic.py", contains code in which initial data analysis was performed to obtain important information. At the end, based on one variable('Elevation'), a simple heuristic logic was written.

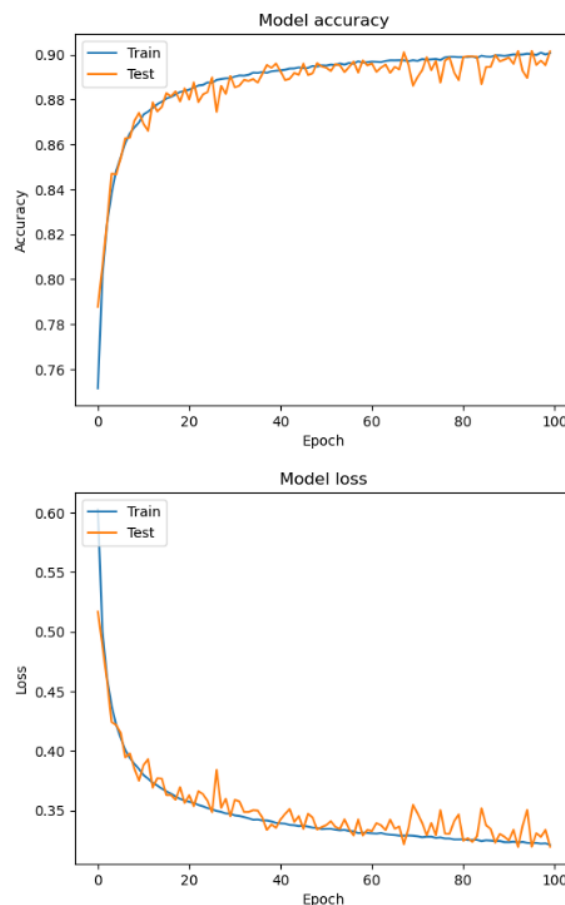Boxplot grouped by Cover_Type

Aspect
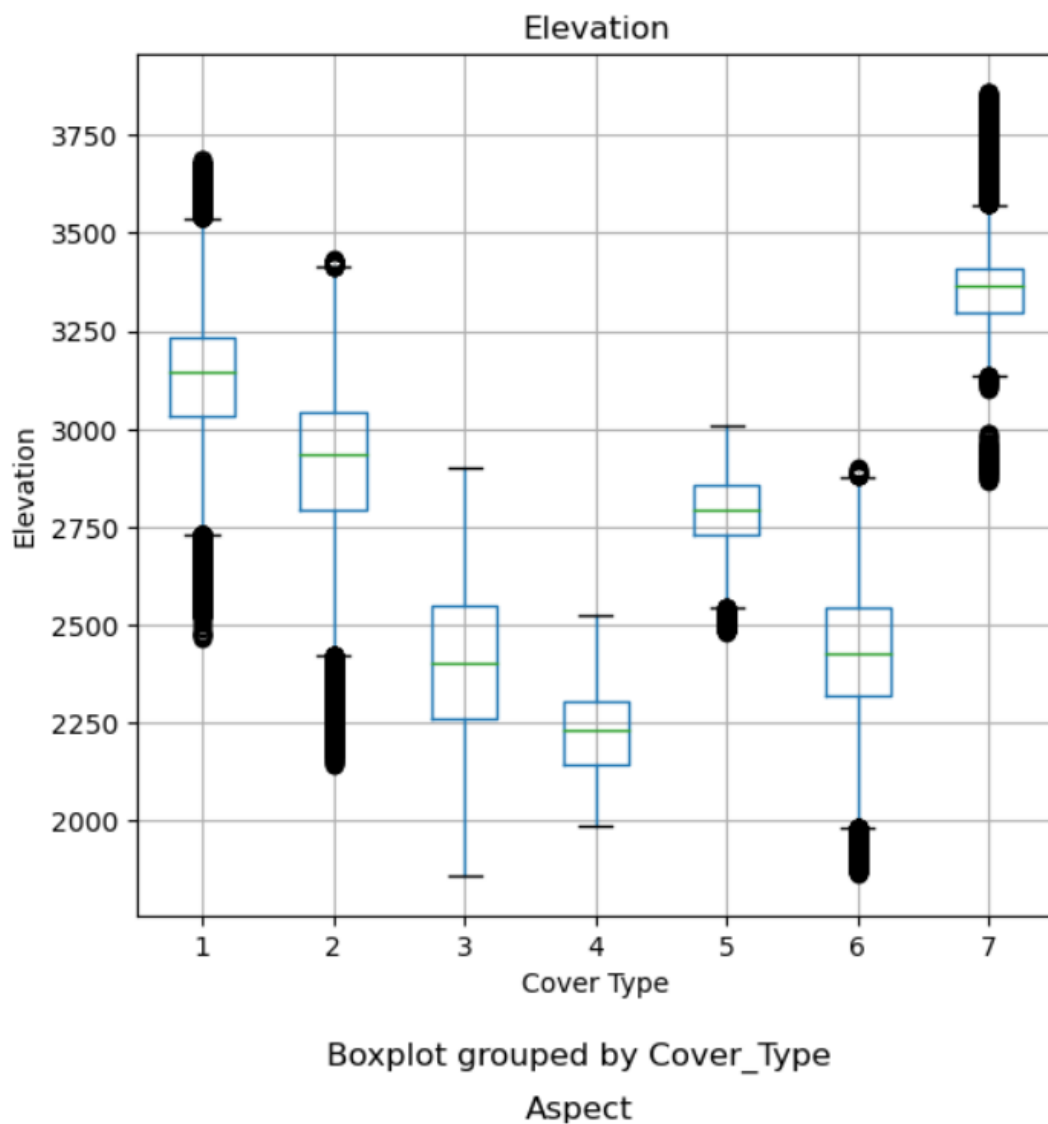
The next file, "Sklearn.py", contains the process of training 2 models: logistic regression and random forest. The third file, named "Tensorflow.py", contains code in which different variants of parameters were tested using functions to find the best one. Then, a custom neural network was created with the best parameters.
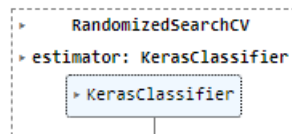

```
print("Best params:", rnd_search_cv.best_params_)
print("Best score:", rnd_search_cv.best_score_)

Best params: {'reg_strength': 0.0001, 'neurons': 128, 'learning_rate': 0.001, 'hidden_layers': 2}
Best score: 0.8922030329704285
```

Next, learning plots for the model were printed, from which it can be inferred that the model was learning properly - the accuracy value increased with each epoch, and the loss function decreased.

Model evaluation was performed for all of these models, and a file called "PlotMetrics.py" was created to visualize the matrices for each model. From them, it can be concluded that in this task, the random forest model performed the best, followed by the neural network, with an accuracy value above 89%.

The next step was to write a simple API that utilized the previously built models. The user was able to choose the model and input data, and then received a prediction using a request. The requests were sent in the following way.

```
Send Request
POST http://localhost:5000/predict/logistic_regression
Content-Type: application/json

{
    "data" : [[2902, 93, 15, 30, 7, 949, 243, 215, 98, 2160, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
}
```

The application is located in the 'rest_api.py' file. To run the script you need to have all model files ('deep_learning.h5', 'LogisticRegression.joblib', 'Random Forest.joblib') in the same folder as the script. Due to too large size, one model has been made available via google drive(link: https://drive.google.com/file/d/1F9NhPMoX6KgiUzJx4SIi0dC4LyuFtfqD/view?usp=sharing)

All files have additional descriptions of used functions.