

Fetal Health Classification Project



The Data:

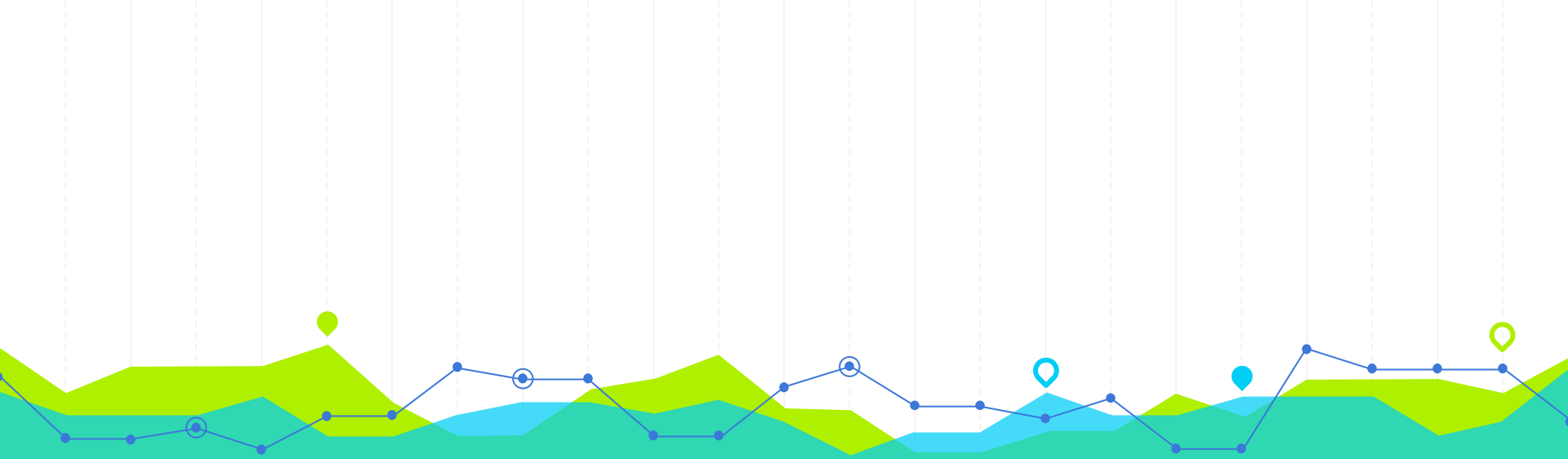
Number of records: 2126

Number of features: 22

Goal: Classify the fetal health into 3 classes:

1. Normal
2. Suspect
3. Pathological

Name: Reema Alnafisi



1 Workflow

Workflow

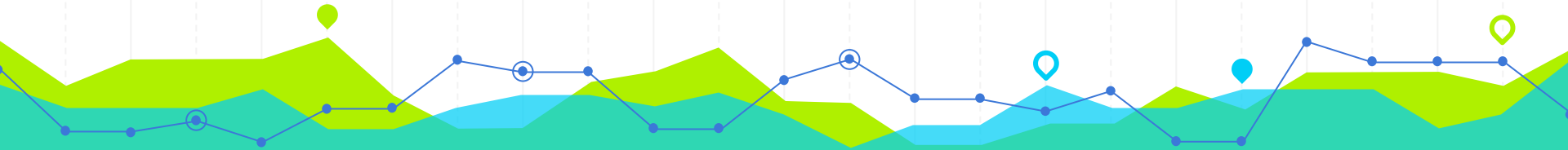
Preprocessing

Data Analysis

Data Scaling

Model
Selection

Model
Building and
Evaluating



2 Preprocessing

Preprocessing

- No null values
- No missing values

Data Preprocessing

How many null values do we have?

```
In [28]: data.isnull().sum().sum()
```

```
Out[28]: 0
```

How many missing values do we have?

```
In [29]: data.isna().sum().sum()
```

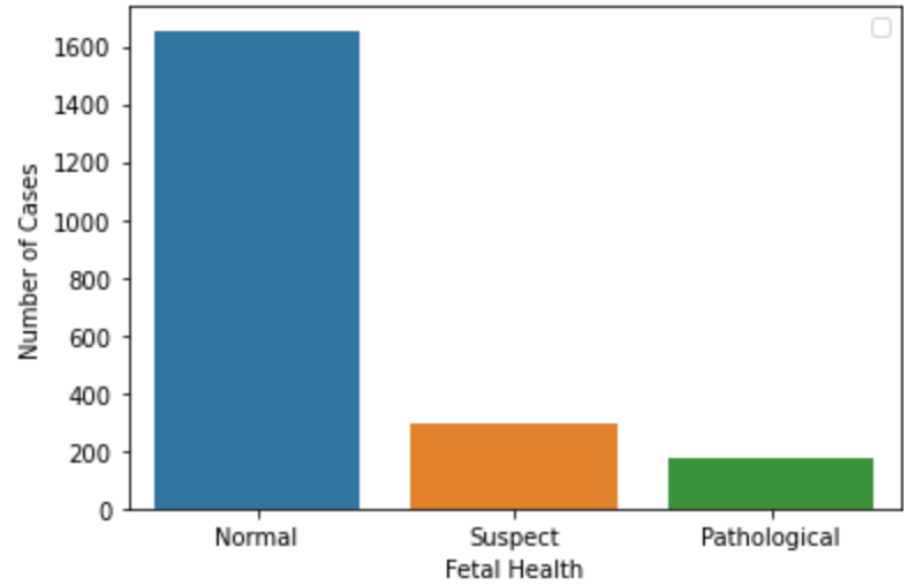
```
Out[29]: 0
```



3 Data Analysis

Data Analysis

Balanced or imbalanced?



Data Analysis

features with largest correlation with the fetal health state are:

- prolonged decelartions
- abnormal short term variability
- percentage of time with abnormal long term variability

fetal_health	
fetal_health	1.000000
prolongued_decelerations	0.484859
abnornal_short_term_variability	0.471191
percentage_of_time_with_abnormal_long_term_variability	0.426146
histogram_variance	0.206630
baseline_value	0.148151
severe_decelerations	0.131934
fetal_movement	0.088010
histogram_min	0.063175
light_decelerations	0.058870
histogram_number_of_zeroes	-0.016682
histogram_number_of_peaks	-0.023666
histogram_max	-0.045265
histogram_width	-0.068789
mean_value_of_short_term_variability	-0.103382



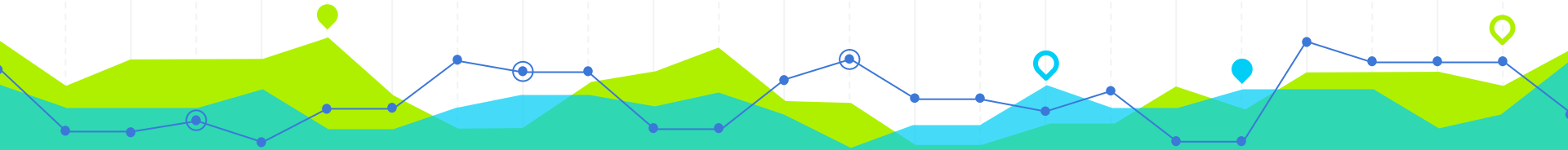
4 Data Scaling

Data Scaling

Feature scaling is a critical step in the preprocessing of data. The most common techniques are:

1. Normalization: used to bound the values between two numbers.
2. Standardization: used to transform the data to have a mean of zero and a variance of one. It makes the data unitless.





5 Model Selection

Model Selection

Cross validation score with 10 folds on a list of pipelines which has different classifiers.

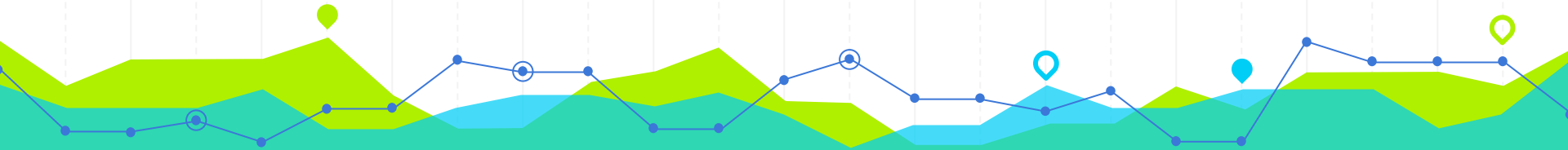
Logistic Regression: 0.897170

Decision Tree: 0.916683

RandomForest: 0.938854

SVC: 0.906594





6 Model Building and Evaluation

Model Building and Evaluation

Random Forest:

An ensemble method, meaning that a random forest model is made up of a large number of small decision trees, called estimators, which each produce their own predictions. The random forest model combines the predictions of the estimators to produce a more accurate prediction.



Model Building and Evaluation

#Find the best estimator according to GridSearchCV

```
best_estimator_RF = CV_RFC.best_params_  
print(f"Best estimator for RF: \n{best_estimator_RF}")
```

Best estimator for RF:

```
{'criterion': 'gini', 'min_samples_leaf': 1, 'min_samples_split': 6, 'n_estimators': 500}
```

#Find the best score according to GridSearchCV

```
best_score_RF = CV_RFC.best_score_  
print(f"Best score for RF: {round(best_score_RF, 4)}")
```

Best score for RF: 0.9388



Model Building and Evaluation

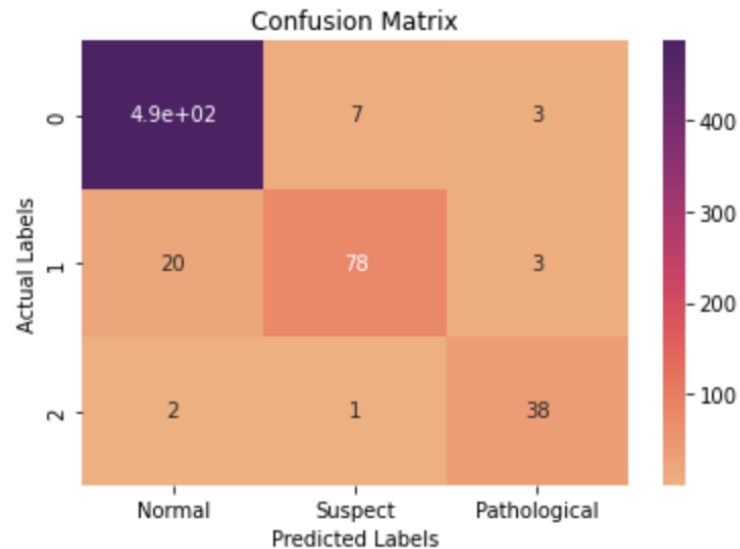
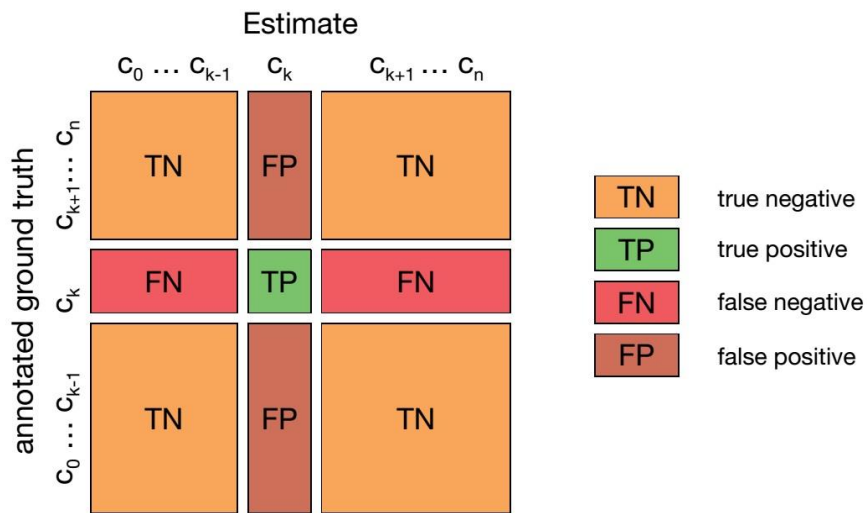
Mean Square Error= 0.08
Root Mean Square Error= 0.283
R² on training set = 0.995
R² on testing set = 0.944

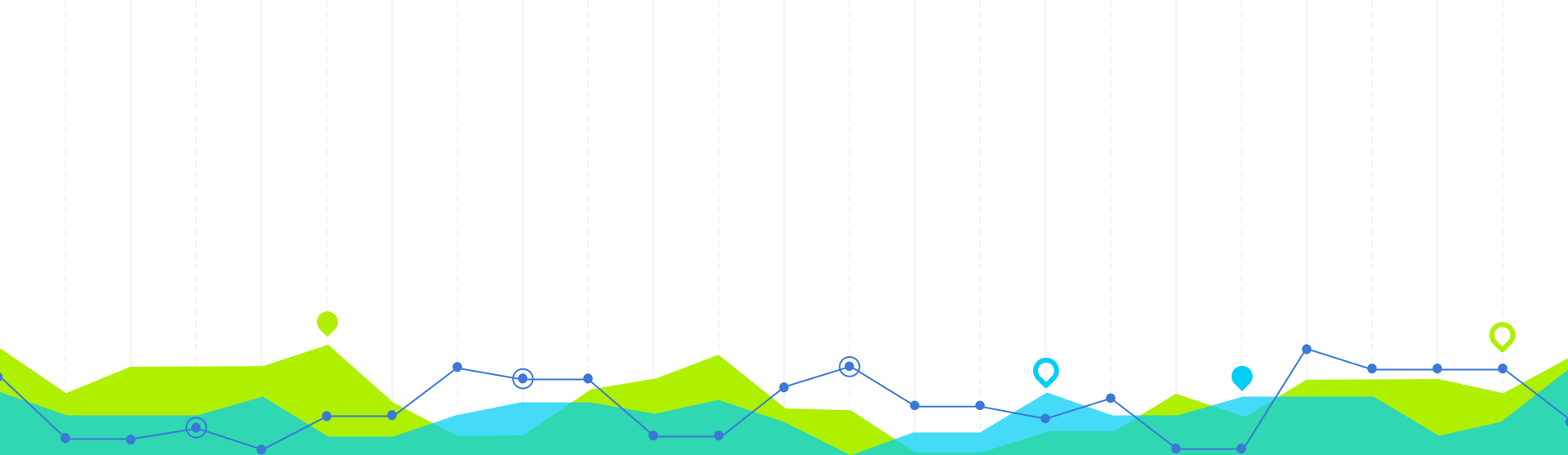
Classification Report

	precision	recall	f1-score	support
1.0	0.96	0.98	0.97	496
2.0	0.91	0.77	0.83	101
3.0	0.86	0.93	0.89	41
accuracy			0.94	638
macro avg	0.91	0.89	0.90	638
weighted avg	0.94	0.94	0.94	638

Model Building and Evaluation

Confusion Matrix:





7 Conclusion

THANKS!

Any questions?

