

Wireframe Documentation

Thyroid Disease Prediction

Contents

1. DataProcessing	3
1.1 Upload Dataset	3
1.2 Handling Missing Values	3
1.3 EDA	4
2. Model Building	4
2.1 Algorithms	5
2.2 Train Models.....	5
3. Model Evaluation	6
3.1 Evaluate Model Metrics.....	Error! Bookmark not defined.
3.2 Compare Algorithms	6

3.3 Most Effective Model for Evaluation.....	7
4. Product Analysis	8
4.1 Accuracy	9
4.2 Future use.....	9
4.3 Future Predictions.....	10

1. Data Processing

```
In [8]: 1 target = df.target
2 create = target.str.split('([A-Za-z]+)', expand=True)
3 create = create[1]
4 target = create.replace({None:'Z'}) #here z is none type
5 df.target = target
```

```
In [9]: 1 df.target.unique()
```

```
Out[9]: array(['Z', 'S', 'F', 'AK', 'R', 'I', 'M', 'N', 'G', 'K', 'A', 'K',
              'J', 'L',
              'MK', 'Q', 'J', 'C', 'O', 'LJ', 'H', 'D', 'GK', 'MI', 'P',
              'FK',
              'B', 'GI', 'GKJ', 'OI', 'E'], dtype=object)
```

1.1 Upload Dataset

First upload the given dataset and remove the unnecessary columns.

```
1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4 import seaborn as sns
5
1 import API_config
2
1 file = open("dataset/thyroid.csv")
2 df = pd.read_csv(file)
3
1 df
```

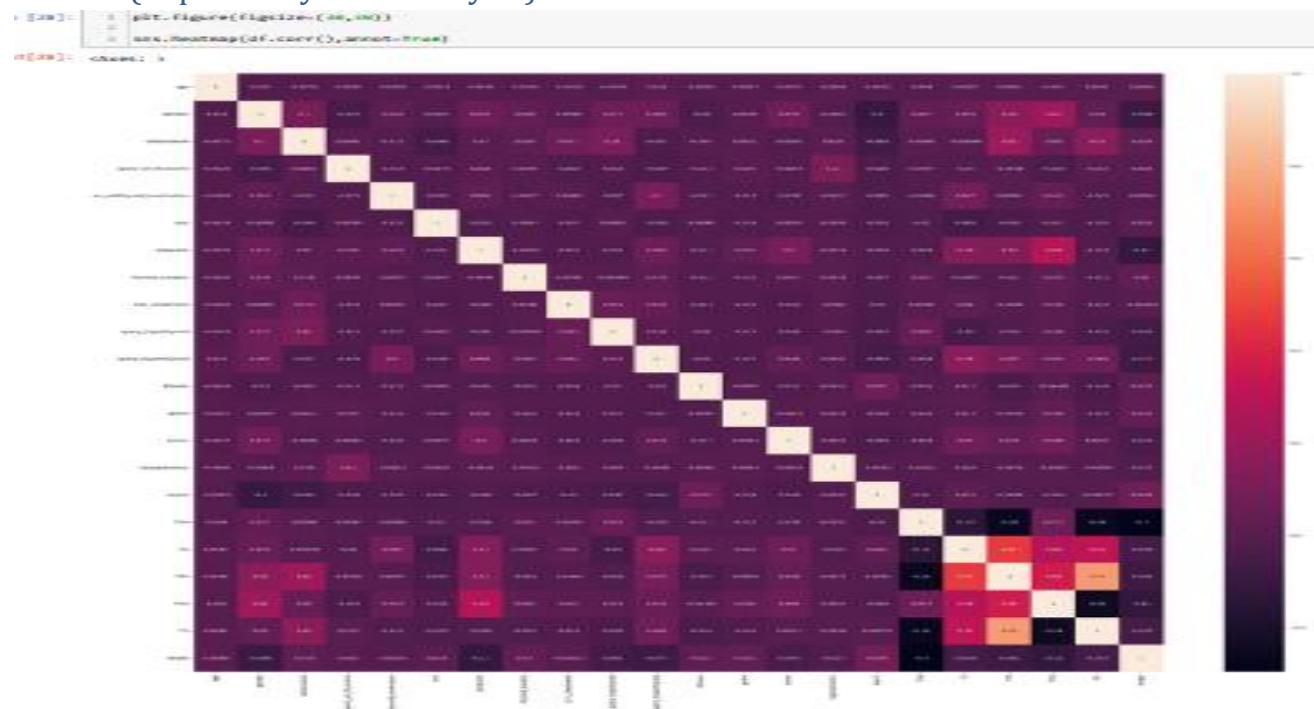
1.2 Handling Missing Values

Missing values are cleared and the unwanted content are removed from the dataset.

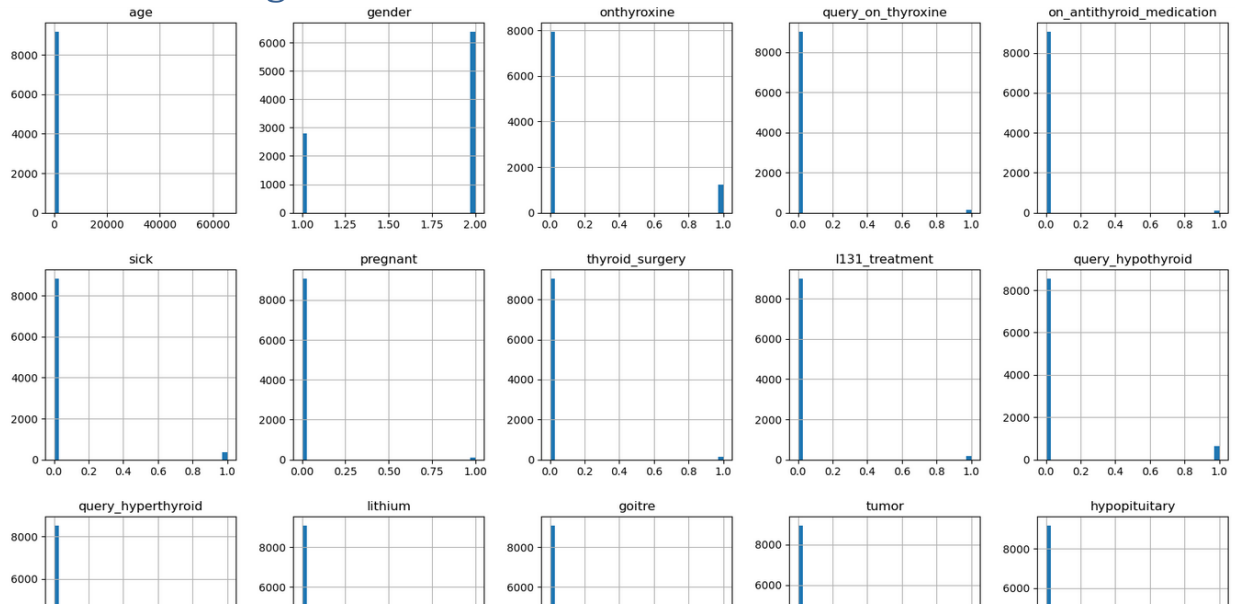
```
5  
4  
5  
6 df = df.replace(['?'],np.nan)
```

```
[12]: 1 df.isnull().sum()
```

1.3 EDA(Exploratory Data Analysis)



2. Model Building



2.1 Algorithms

1. Decision tree
2. Random Forest
3. K-NN Classifier
4. SVM
5. Logistic Regression

2.2 Train Models

SVM

```
[ ]: 1 from sklearn.svm import SVC
      2 svm = SVC(kernel="sigmoid")
      3 sclf = svm.fit(X_train,y_train)
      4 y_pred = sclf.predict(X_test)
      5
```

```
[ ]: 1 accuracy_score(y_pred,y_test)
```

```
[ ]: 0.727803738317757
```

logisitic Regression

```
[ ]: 1 from sklearn.linear_model import LogisticRegression
      2 lr = LogisticRegression(max_iter=1000)
      3 lrclf = lr.fit(X_train,y_train)
      4 y_pred = lrclf.predict(X_test)
      5 accuracy_score(y_pred,y_test)
```

```
[ ]: 0.7492211838006231
```

3.Model Evaluation

```
[45]: 1 from sklearn.metrics import accuracy_score
```

3.1 Evaluation Metrics

KNN Classifier

```
In [50]: 1 accuracy_score(y_pred,y_test)
```

```
Out[50]: 0.8419003115264797
```

Decision Tree and Random Forest

Decisiontree

```
In [46]: 1 from sklearn.tree import DecisionTreeClassifier
          2 tree = DecisionTreeClassifier(max_depth=3)
          3 clf = tree.fit(X_train,y_train)
          4 treepredict = clf.predict(X_test)
```

```
In [47]: 1 accuracy_score(treepredict,y_test)
```

```
Out[47]: 0.7975077881619937
```

Random Forest

```
In [48]: 1 from sklearn.ensemble import RandomForestClassifier
          2 rf = RandomForestClassifier(max_depth=2,n_estimators=200)
          3 rclf = rf.fit(X_train,y_train)
          4 rfpred = rclf.predict(X_test)
          5 accuracy_score(rfpred,y_test)
```

```
Out[48]: 0.742601246105919
```

3.2 Compare Algorithms(With PCA-Accuracy may Differ)

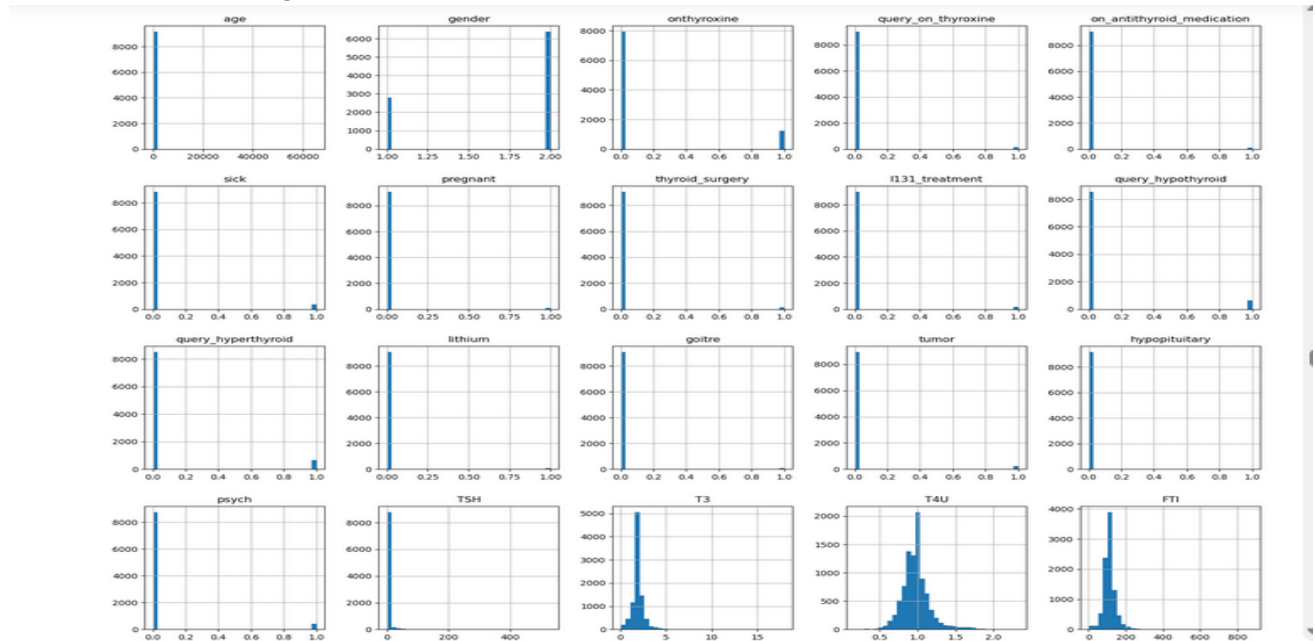
Model	Accuracy
Decision Tree	0.8430685358255452
Random Forest Classifier	0.7387071651090342
KNN Classifier	0.7387071651090342
SVM	0.7422118380062306

3.3 Top Model for Evaluation

```
In [50]: 1 accuracy_score(y_pred,y_test)
```

```
Out[50]: 0.8419003115264797
```

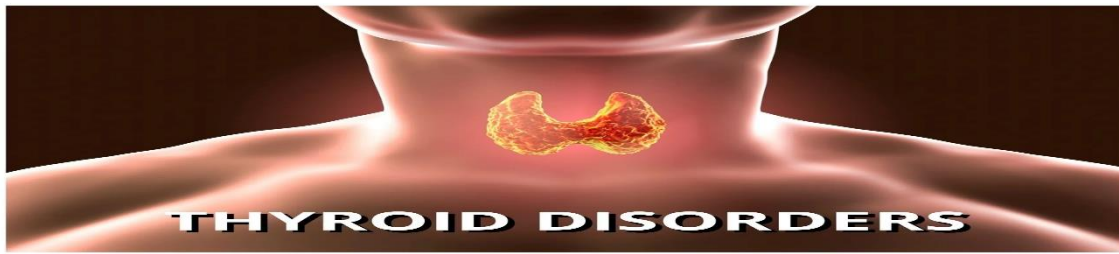
4. Product Analysis



4.1 Accuracy

True Label Class 1	Correct True Positive for Class 1	Wrong False Positive for Class 2
	Wrong False Positive for Class 1	Correct True Positive for Class 2
True Label Class 2	Predicted Label Class 1	Predicted Label Class 2

4.2 Future Use



Cured At Early Stages

4.3 Future Prediction

Factors that inhibit proper production of thyroid hormones

- Stress
- Infection, trauma, radiation, medications
- Fluoride (antagonist to iodine)
- Toxins: pesticides, mercury, cadmium, lead
- Autoimmune disease: Celiac

Factors that increase conversion of T4 to RT3

- Stress
- Trauma
- Low-calorie diet
- Inflammation (cytokines, etc.)
- Toxins
- Infections
- Liver/kidney dysfunction
- Certain medications

Factors that contribute to proper production of thyroid hormones

- Nutrients: iron, iodine, tyrosine, zinc, selenium
- vitamin E, B2, B3, B6, C, D

Factors that increase conversion of T4 to T3

- Selenium
- Zinc

Factors that improve cellular sensitivity to thyroid hormones

- Vitamin A
- Exercise
- Zinc

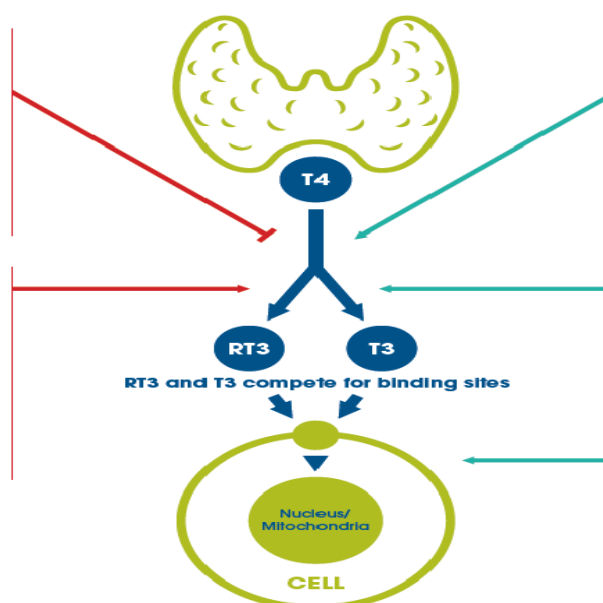


Figure 1 Factors that Affect Thyroid Function (The Institute for Functional Medicine)