Wireframe Documentation
Thyroid Disease Prediction

Contents

1. Data Processing

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

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

Out[9]:
```
array(['Z', 'S', 'F', 'AK', 'R', 'I', 'M', 'N', 'G', 'K', 'A', 'KJ', '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.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
import API_config
```

```
file = open("dataset/thyroid.csv")
df = pd.read_csv(file)
```

```
df
```

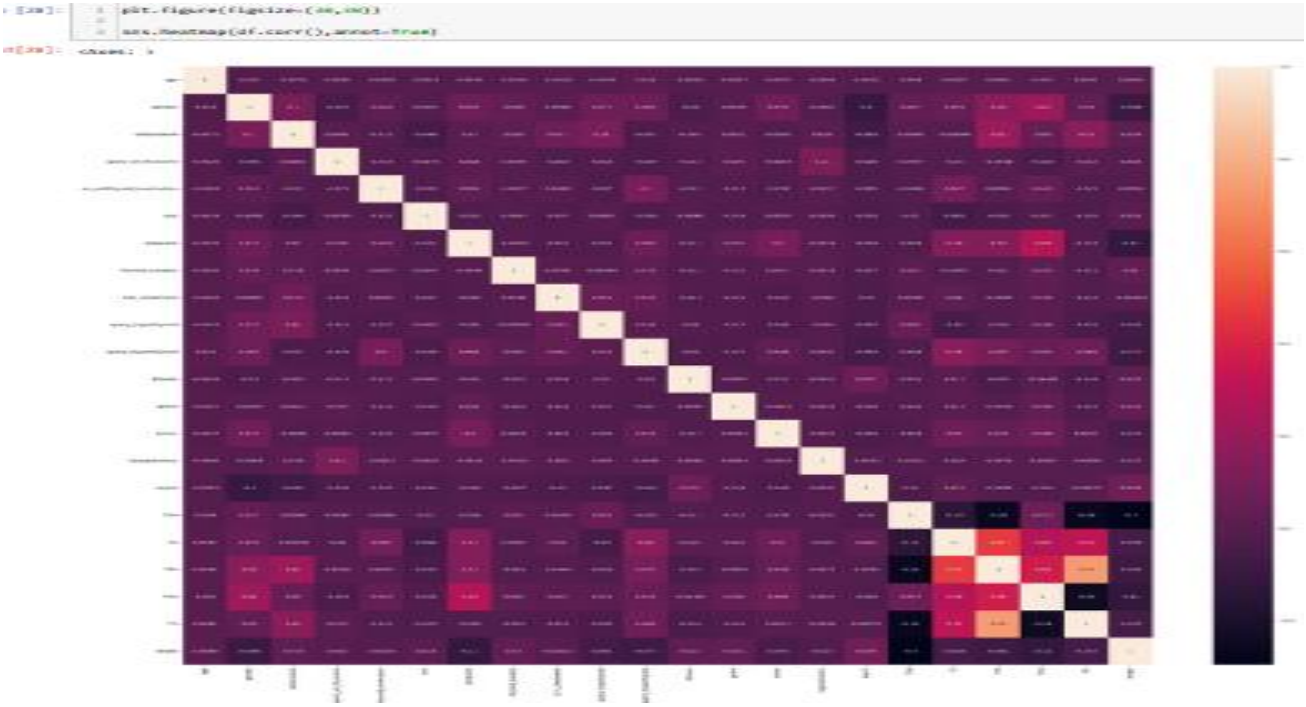
1.2 Handling Missing Values

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

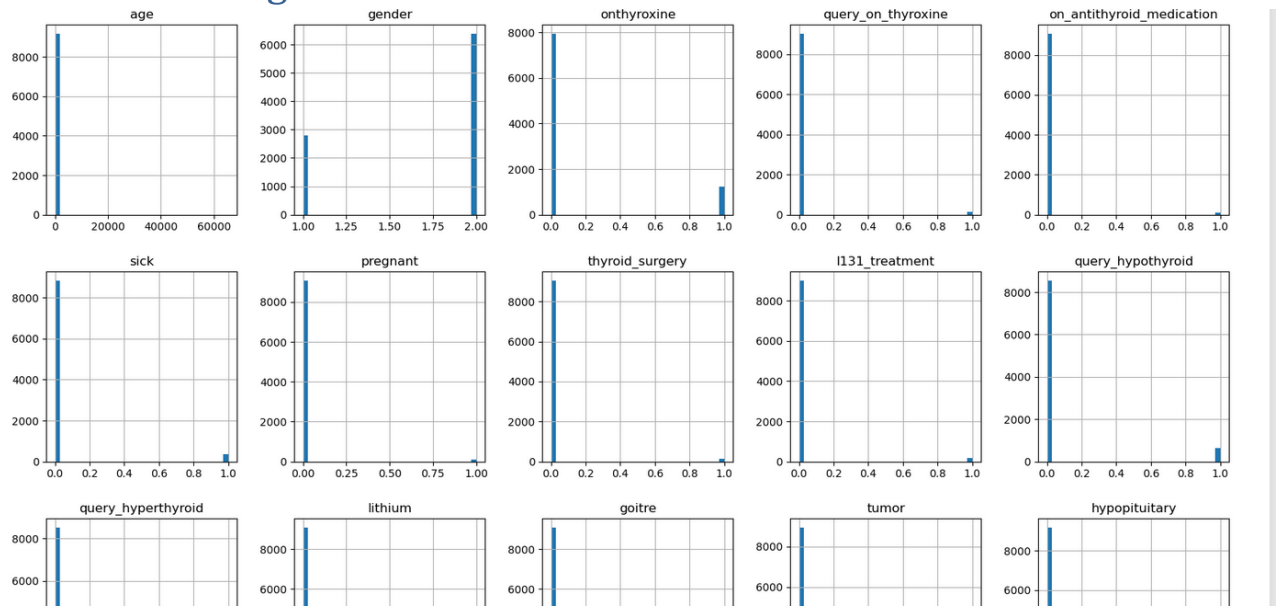
```
df = df.replace(['?'],np.nan)
```

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

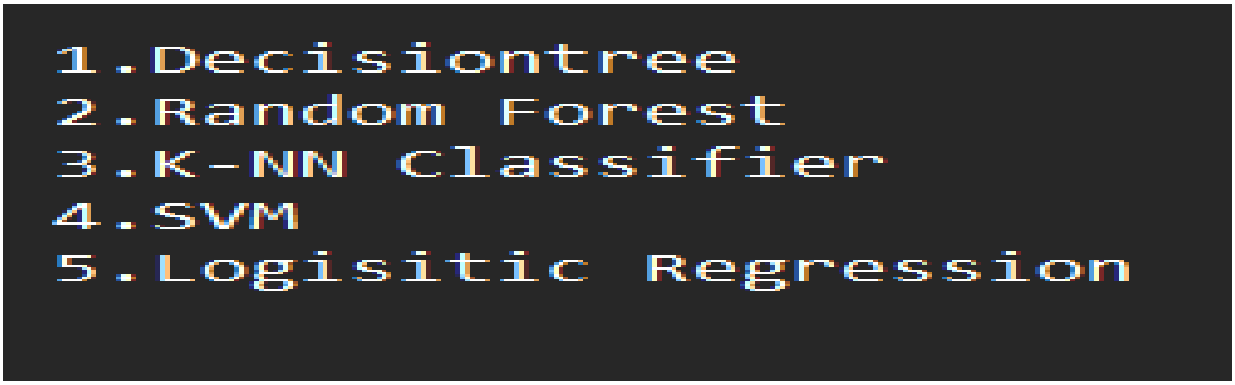
1.3 EDA(Exploratory Data Analysis)



2. Model Building



2.1 Algorithms



2.2 Train Models

SVM

```
In [ ]:
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

In [ ]:
1 accuracy_score(y_pred,y_test)

Out[ ]: 0.727803738317757
```

logisitic Regression

```
In [ ]:
1 from sklearn.linear_model import LogisticR
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)

Out[ ]: 0.7492211838006231
```

3.Model Evaluation

```
In [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)

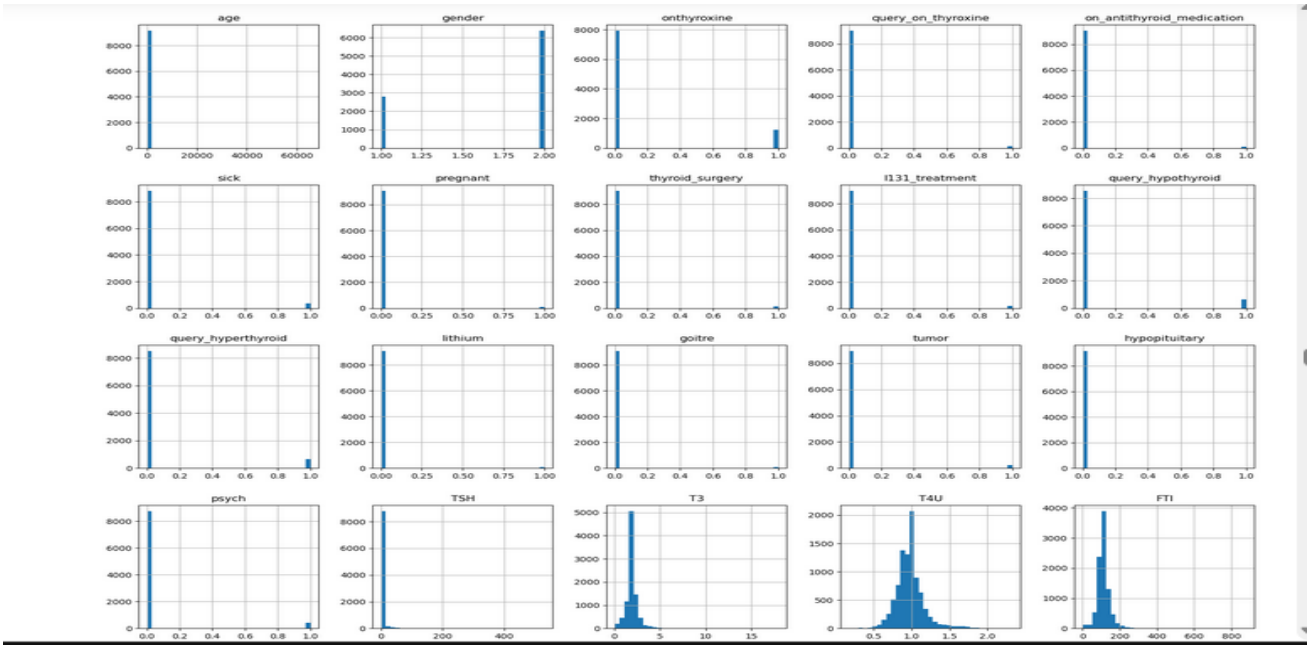
Model	Accuracy
Decision Tree	0.8430685358255452
Random Forest Classifier	0.7387071651090342
KNN Classifier	0.7387071651090342
SVM	0.7422118380062306

3.3 Top Model for Evaluation

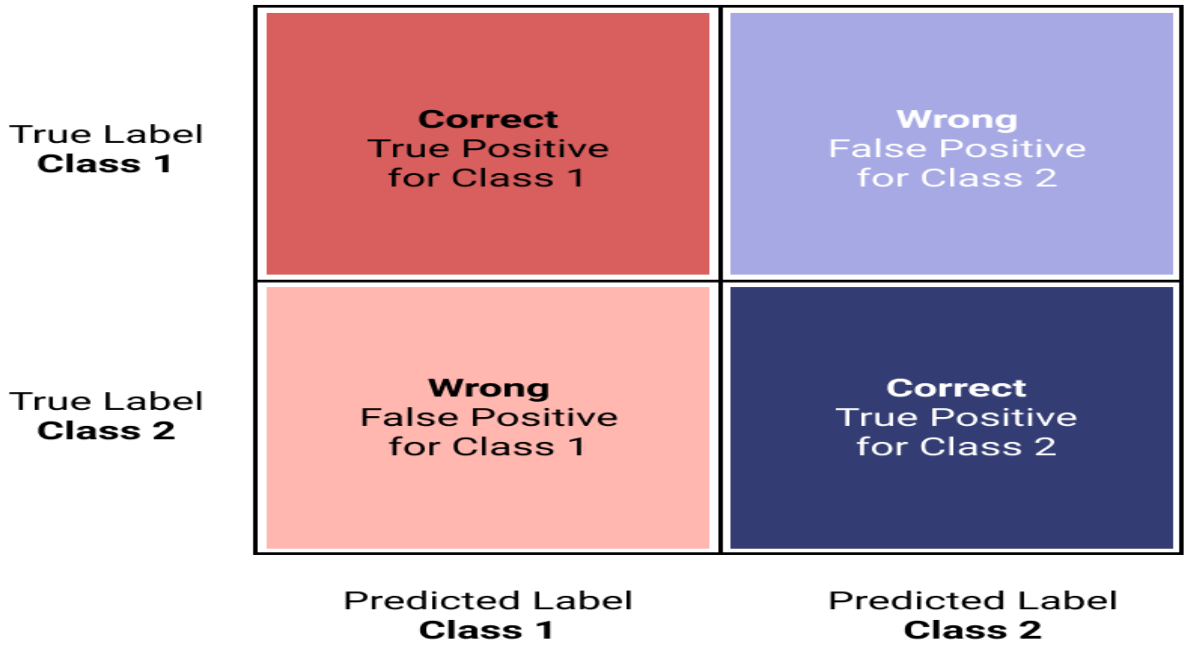
```
In [50]:  1  accuracy_score(y_pred,y_test)

Out[50]: 0.8419003115264797
```

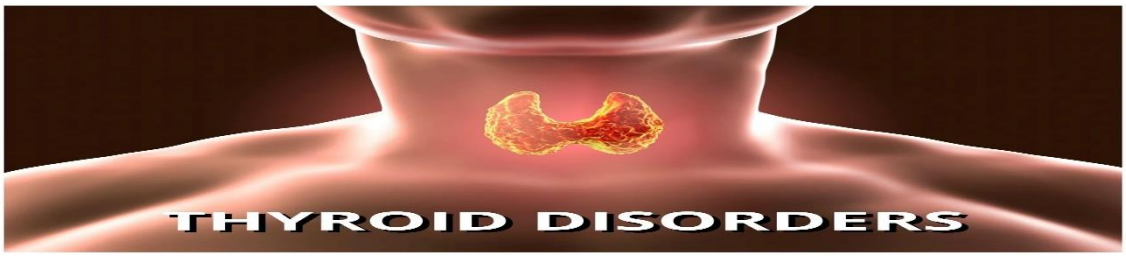
4. Product Analysis



4.1 Accuracy

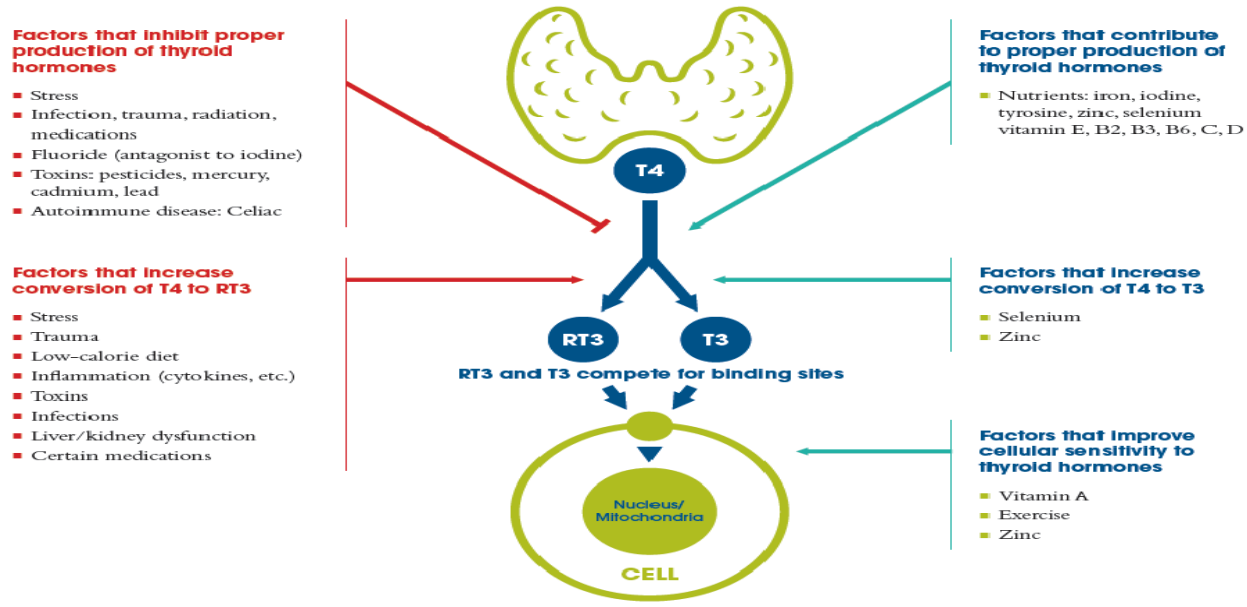


4.2 Future Use



Cured At Early Stages

4.3 Future Prediction



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