Thyroid Disease Prediction (thyroid.ipynb)

- **Problem Statement**: The project aims to predict whether a patient has thyroid disease (target classes likely 'negative', 'hyperthyroid', 'hypothyroid', etc., making it a multi-class classification problem).
- **Dataset Used**
  :
  - The dataset was loaded from hypothyroid.csv.
  - The target variable is 'binaryClass' (P/N), making this a binary classification problem as implemented, although thyroid disease can have more states. The original target in the CSV seems to be the last column which might be a more detailed class, but the notebook focuses on 'binaryClass'.

- **Methodology and Approach**
  :
  - **Data Preprocessing**
    :
    - Numerous columns with excessive missing values (more than 200 NaNs) were dropped.
    - Remaining rows with any missing values were dropped using dropna().
    - Categorical features (like 'sex', 'on_thyroxine', 'query_on_thyroxine', etc., and the target 'binaryClass') were converted to numerical representations using LabelEncoder.
    - Numerical features were scaled using StandardScaler.

  - **Model Training**
    :
    - The data was split into training (75%) and testing (25%) sets.
    - The following classification models were implemented:
      - Logistic Regression (LogisticRegression)
      - K-Nearest Neighbors (KNeighborsClassifier)
      - Support Vector Classifier (SVC)
      - Decision Tree Classifier (DecisionTreeClassifier)
      - Random Forest Classifier (RandomForestClassifier)
      - Gaussian Naive Bayes (GaussianNB)

- **Model Evaluation**
  :
  - Models were evaluated using accuracy scores, confusion matrices, and classification reports.

- **Results and Conclusion**
  :
  - Logistic Regression: Accuracy 97.65%.
  - K-Nearest Neighbors: Accuracy 97.52%.
  - Support Vector Classifier: Accuracy 97.65%.
  - Decision Tree: Accuracy 97.78%.
  - Random Forest: Accuracy 98.16%.
  - Gaussian Naive Bayes: Accuracy 94.74%.
  - The Random Forest Classifier achieved the highest accuracy.
  - The Random Forest model was saved to thyroid_model.pkl.