


Capstone Project

Project Documentation

Forensic Dentistry: Gender Prediction Using Dental Metrics

 **Project Title:** Using Dental Metrics to Predict Gender

 **Author:** Saurabh Sharma

 **Date:** 01-04-2025

Introduction

◆ Problem Statement

Forensic dentistry is a branch of forensic medicine that plays a crucial role in identifying individuals, especially in cases where bodies are unrecognizable due to natural disasters or other incidents. Dental structures, particularly **teeth**, are highly durable and can serve as reliable indicators of identity. This project leverages **machine learning** to predict an individual's **gender** based on dental measurements.

◆ Objective

- To analyze dental metrics and their relationship with gender.
 - To implement **machine learning models** for gender classification.
 - To evaluate the performance of different classifiers and identify the best-performing model.
-

Dataset Description

◆ Dataset Name: Dentistry Dataset.csv

This dataset contains various dental measurements, including inter-canine distance, canine width, and canine index, used to classify gender.

◆ Key Features:

Feature Name	Description
Age	The age of the individual
Gender (Target Variable)	Male (1) / Female (0)
Sample ID & SL No.	Unique identifier (not used for prediction)
Inter-canine distance intraoral	Measurement between upper canine teeth
Right & Left Canine Width Casts	Width of the right and left canines
Canine Index	Canine index measurement

Capstone Project

◆ **Target Variable:**

- **Gender** (Male = 1, Female = 0)
 - **Independent Variables:** Various dental measurements
-

3 **Methodology**

◆ **Step 1: Data Preprocessing**

- Handle **missing values** (if any).
- Encode categorical variables (**Label Encoding** for Gender).
- Drop unnecessary columns (**Sample ID, SL No.**).
- Normalize numerical features (**Scaler or Normalizer**).

◆ **Step 2: Exploratory Data Analysis (EDA)**

- **Heatmap to check feature correlations**
- **Histograms to visualize distributions of features**

◆ **Step 3: Model Building**

- Split data into **training (80%) and testing (20%)**.
- Train the following classification models:
 1. **Logistic Regression**
 2. **Decision Tree**
 3. **Random Forest**
 4. **XGBoost**

◆ **Step 4: Model Evaluation**

- **Accuracy Score**
 - **Confusion Matrix**
 - **ROC-AUC Curve**
-

4 **Results & Analysis**

◆ **Model Performance Comparison**

Model	Accuracy (%)
Logistic Regression	0.64090909090909
Decision Tree	0.87727272727272

Capstone Project

Model	Accuracy (%)
Random Forest	0.895454545454545
XGBoost	0.9

🔑 Key Findings:

- The best model, **XGBoost**, achieved an accuracy of **XX%**.
- The **heatmap analysis** showed that certain **dental features are strongly correlated with gender**.

5 Conclusion & Future Work

✅ Conclusion

- This project successfully built a **gender classification model** using dental metrics.
- The best model achieved an **accuracy of XX%**, demonstrating potential for forensic applications.

🔮 Future Work

- **Expand dataset:** More diverse samples for better accuracy.
- **Use Deep Learning:** Implement Neural Networks for improved classification.
- **Test on real-world forensic datasets:** Validate the model in forensic investigations.

6 References

- **Scikit-learn Documentation:** <https://scikit-learn.org>
- **XGBoost Documentation:** <https://xgboost.readthedocs.io>
- **Research Papers on Forensic Dentistry**

📦 Installation & Usage

💎 Clone the Repository

git clone <https://github.com/Saurabhji-1/Capstone-Project/tree/main>

💎 Install Dependencies

pip install -r requirements.txt

💎 Run the Jupyter Notebook

jupyter notebook