

Project Documentation

Project Title

Using Dental Metrics to Predict Gender

Abstract

This project focuses on leveraging dental measurements to predict an individual's gender. By utilizing various machine learning models and exploring relationships among features, we aim to identify patterns that allow accurate classification. The project is rooted in healthcare and forensic science, providing a practical approach to gender estimation based on dental metrics.

Objectives

- Analyze the dataset containing dental metrics.
- Preprocess the data for optimal model performance.
- Conduct Exploratory Data Analysis (EDA) to discover meaningful insights.
- Build and evaluate multiple classification models to predict gender.
- Determine the best-performing model based on evaluation metrics.

Tools & Technologies Used

- Programming Language: Python
- Tools: Jupyter Notebook, MS Excel
- Libraries: pandas, numpy, seaborn, matplotlib, scikit-learn, xgboost

Domain

Healthcare & Forensic Science

Dataset Description

- The dataset comprises 1100 rows and 14 columns.
- It includes various dental measurements and the target column: Gender.

- Features include measurements such as left/right canine width, incisor width, and other tooth dimensions.

Methodology

- Import Libraries & Data: Necessary Python libraries are imported, and data is loaded into a DataFrame.
- Data Preprocessing: Includes handling null values, encoding categorical variables, and normalization.
- EDA (Exploratory Data Analysis): Correlation matrix and visualizations revealed a strong positive correlation between 'left canine width (casts)' and 'Gender' (correlation coefficient: 0.64).
- Model Building:
 - Models Trained: Logistic Regression, Decision Tree, Random Forest, XGBoost Classifier.
 - Best Model: Decision Tree Classifier, based on accuracy and interpretability.
- Evaluation:
 - Used metrics like Accuracy, Confusion Matrix, ROC Curve to compare models.

Results

- Decision Tree Classifier outperformed other models.
- Strong feature importance was observed for left canine width.
- Model exhibited reliable performance in gender classification.

Conclusion

The analysis demonstrated that dental metrics, particularly left canine width, serve as strong indicators for gender prediction. The Decision Tree model emerged as the most effective model among those tested.

Future Work

- Incorporate more dental features or cross-sectional data.
- Expand dataset size and diversity for improved generalization.
- Explore deep learning approaches for better accuracy and feature extraction.

References

- Dataset source : Dentistry Dataset.csv

- Scikit-learn, XGBoost documentation

- Research papers on forensic odontology (

- 1) "Mandibular and Dental Measurements for Sex Determination Using Machine Learning"

- 2) "Sex Determination in Forensic Odontology: A Review")