**Capstone Project Documentation**

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

**Objective**

The objective of this project is to analyze dental metrics and predict the gender of an individual using machine learning techniques. This can be useful in forensic dentistry to identify deceased individuals when other identification methods are unavailable.

**Background & Scope**

Forensic dentistry is a branch of forensic medicine that aids in identifying deceased individuals based on dental measurements. Teeth are durable and can withstand environmental factors, making them ideal for identification. This project aims to use dental measurements to determine gender, helping forensic experts during natural calamities or accidents.

**Project Architecture**

1. **Raw Data Collection:** Collecting measurements from dental samples.
2. **Data Preprocessing:** Handling missing values, encoding categorical data, and normalizing features.
3. **Exploratory Data Analysis (EDA):** Analyzing correlations using heatmaps.
4. **Model Building:** Using classification algorithms to build predictive models.
5. **Evaluation:** Evaluating models using metrics like accuracy, ROC curves, and AUC.
6. **Reporting:** Documenting results and conclusions.

**Dataset Information**

* **Age:** The person's age in years
* **Gender:** Target variable (Male/Female)
* **SampleID & SL No.:** Unique identifiers for individuals
* **Inter-Canine Distance (Intraoral/Casts), Right & Left Canine Cast:** Measurements used as independent variables.

**Step-by-Step Implementation**

1. **Data Collection:**
   * Load dataset using Pandas.
   * Inspect dataset for missing values and outliers.
2. **Preprocessing:**
   * Handle missing values appropriately.
   * Encode categorical variables using LabelEncoder.
   * Normalize features using Normalizer from sklearn.
3. **EDA:**
   * Analyze feature relationships using heatmaps.
   * Identify multicollinearity and drop unimportant features.
4. **Model Building:**
   * Train/Test Split using train\_test\_split from sklearn.
   * Models used: Logistic Regression, Decision Tree, Random Forest, XGBoost.
5. **Evaluation:**
   * Evaluate using accuracy, confusion matrix, ROC, and AUC curves.

**Results & Conclusion**

* The model with the highest accuracy was **Random Forest**, delivering the most reliable results for gender prediction.
* This approach provides an efficient method for forensic identification, proving the feasibility of dental metrics in gender prediction.