

Problem Identification

Problem Statement Formation:

The problem at hand is to investigate trends and patterns in in-vitro fertilization (IVF) success rates over time, specifically focusing on differences between fresh and vitrified-warmed cycles. Additionally, the goal is to explore factors that may influence clinical pregnancy rates and identify potential areas for improvement in fertility treatment protocols.

Context:

The fertility industry continuously seeks to enhance the effectiveness of IVF treatments by understanding the factors contributing to success rates and identifying areas for optimization. Analyzing trends and patterns in IVF success rates and identifying influential factors can provide valuable insights for clinicians and researchers in the field.

Criteria for Success:

The success of this project will be determined by the ability to:

- Identify significant trends and patterns in IVF success rates over time.
- Differentiate between the success rates of fresh and vitrified-warmed cycles.
- Uncover factors that significantly influence clinical pregnancy rates.
- Provide actionable recommendations for improving fertility treatment protocols based on the analysis results.

Scope of Solution Space:

The solution space encompasses the analysis of historical IVF data to understand trends, patterns, and factors influencing clinical pregnancy rates. It includes data preprocessing, exploratory data analysis (EDA), statistical analysis, predictive modeling, feature importance analysis, and deriving actionable insights and recommendations.

Constraints:

- **Availability and quality of data:** Limited access to comprehensive datasets containing relevant patient information and treatment outcomes may constrain the development of the predictive model.

Stakeholders:

- **Fertility clinics and healthcare providers:** Interest in improving IVF success rates and optimizing treatment plans for their patients.
- **Patients undergoing fertility treatments:** Seek accurate predictions of IVF success to make informed decisions about treatment options and expectations.
- **Researchers and academia:** Benefit from advancements in predictive modeling techniques and insights into factors influencing IVF outcomes.

Data Sources:

- The primary data source for this project is the provided dataset containing information on fresh and vitrified-warmed IVF cycles, including the number of ovarian pick-up (OPU) procedures, embryo transfers, and clinical pregnancy rates. Additionally, supplementary data sources will be added that have less aggregated data.

Approach:

1. **Data Cleaning and Preparation:**
 - a. Convert the provided data into a structured format suitable for analysis, handling missing or inconsistent values, and ensuring proper data types.
2. **Exploratory Data Analysis (EDA):**
 - a. Visualize trends in OPU procedures, embryo transfers, and clinical pregnancy rates over time.
 - b. Compare clinical pregnancy rates between fresh and vitrified-warmed cycles and analyze seasonal variations or trends.
3. **Statistical Analysis:**
 - a. Conduct statistical tests to compare mean clinical pregnancy rates between fresh and vitrified-warmed cycles and explore correlations with other variables.
4. **Predictive Modeling:**
 - a. Build predictive models to forecast clinical pregnancy rates based on relevant factors, evaluating model performance and identifying key predictors of IVF success.
5. **Feature Importance Analysis:**
 - a. Determine the importance of different factors in predicting IVF success using feature importance techniques.
6. **Recommendations and Insights:**

- a. Provide insights into factors influencing IVF success rates and suggest adjustments to treatment protocols or patient selection criteria based on the analysis results.

7. Deliverables:

- a. GitHub repository containing:
 - i. Code for data preprocessing, modeling, and evaluation.
 - ii. Project report documenting the methodology, results, and conclusions.
 - iii. Slide deck summarizing the project findings and insights.