Future Baby Names Prediction: Leveraging Famous Influences with Machine Learning

Machine Learning Models Analysis

YING ZHOU

09/16/2024

Contents

Introduction

A baby's name is a significant gift from parents, embodying their hopes, love, and aspirations for their child.

The process of naming a child is deeply influenced by social experiences, cultural sensitivities, and historical trends.

This project aims to predict popular baby names using machine learning by analyzing historical data from various sources.

Research Value

Cultural Reflection: Names often reflect cultural, social, and generational identities, contributing to a child's sense of belonging.

Identity and Aspiration: Parents often choose names that align with their aspirations for their children, influencing future generations.

Market Insights: Companies in baby products, toys, and educational services can benefit from understanding naming trends.

Dataset

Basic Features:

- Year
- Gender_Binary
- Name

Name Features:

- Name_Length
- Vowel_Count
- Consonant_Count
- Vowel_Ratio
- Ends_With_Specified_Letters

Other Features:

- Year_of_Last_Appearance
- Is_Famous
- Is_Top_100

Dataset

Count Features:

- Count
- Name_Ratio
- Gender_Name_Ratio

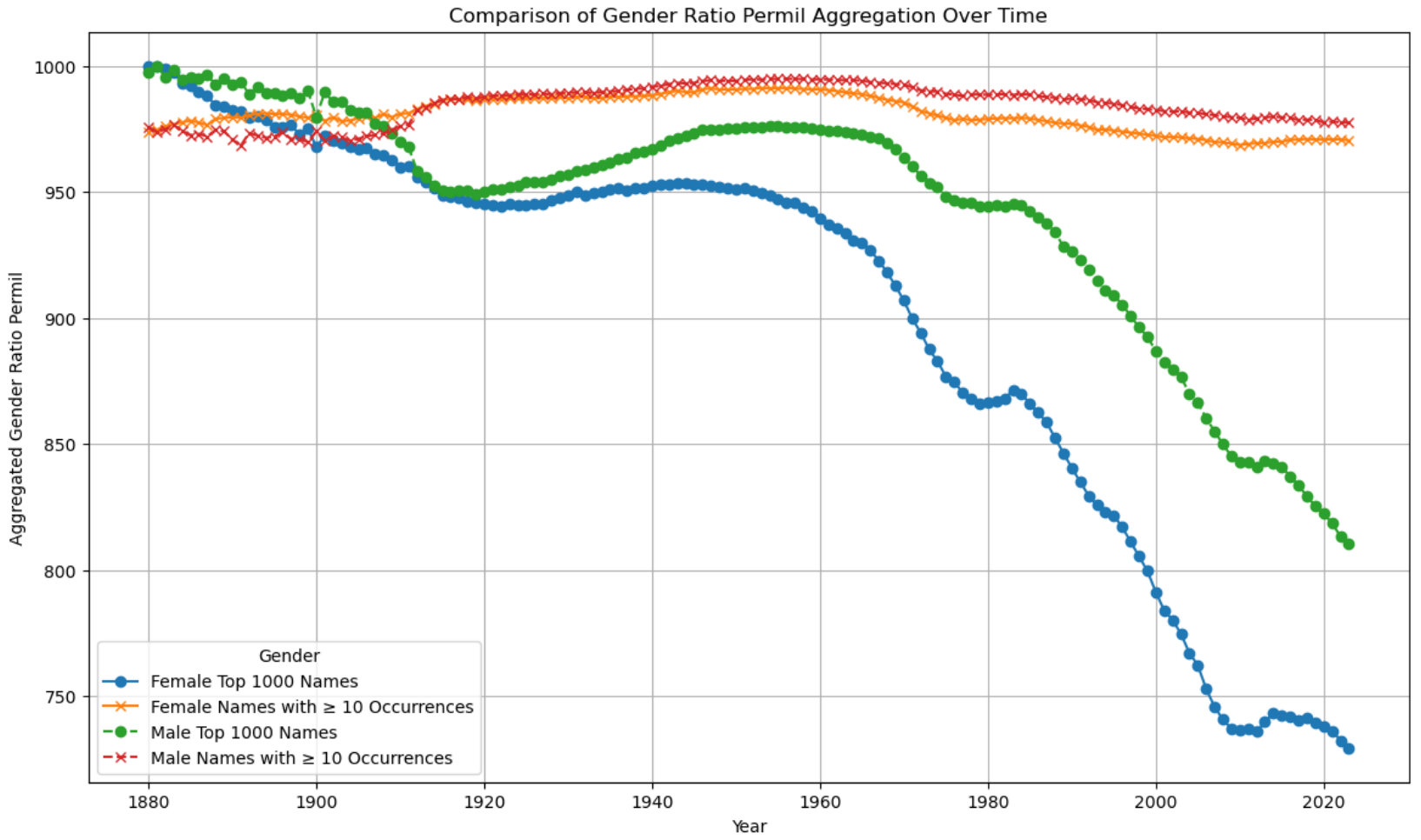
Yearly Changes Features:

- Yearly_Change_Count
- Yearly_Change_National_Ratio
- Yearly_Change_Gender_Ratio

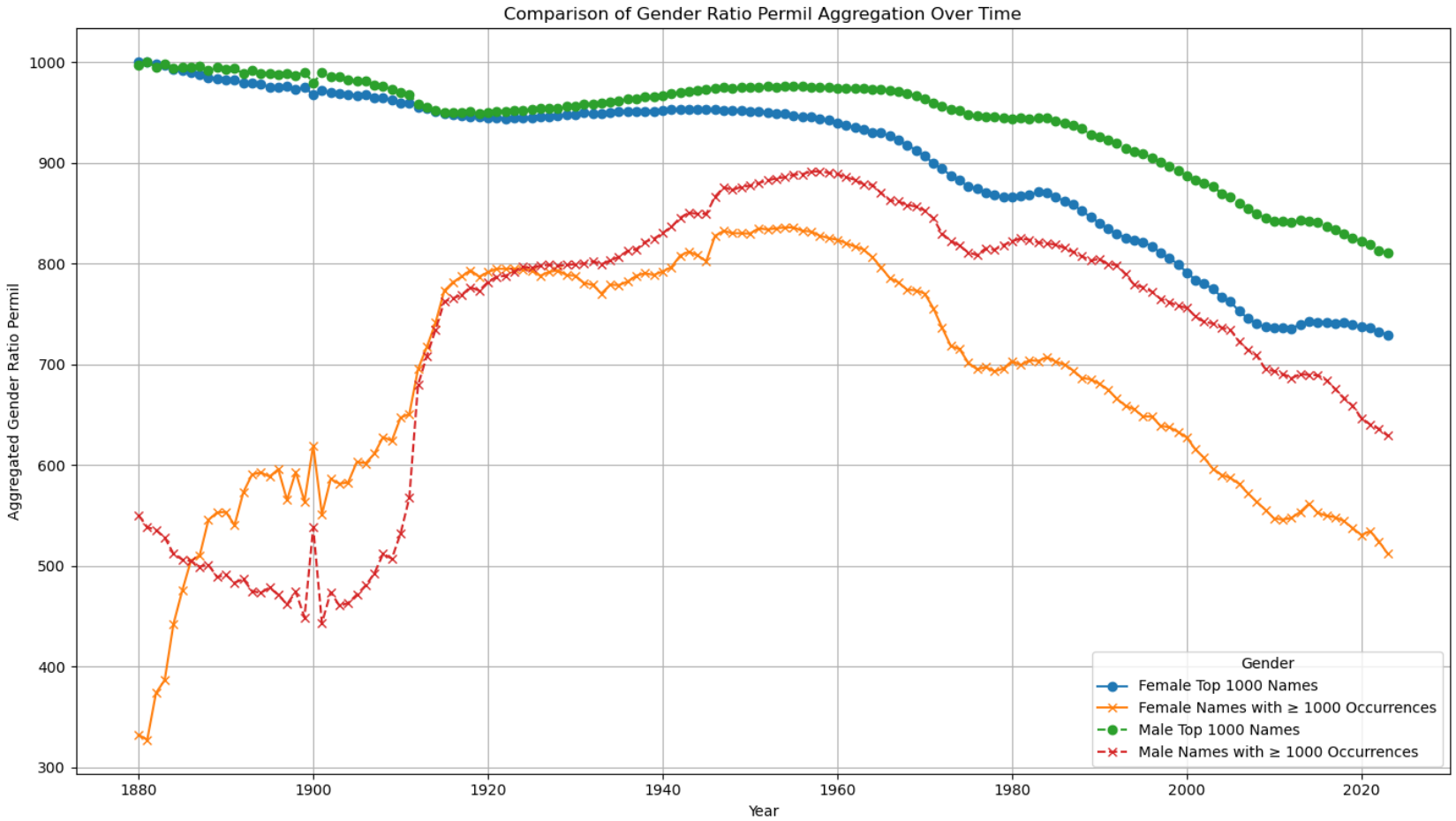
5 Year Window Features:

- Rolling_Average_Count_5_Years
- Rolling_Average_Gender_Ratio_5_Years
- Rolling_Average_National_Ratio_5_Years

Top 1000 names vs. Name Occurances More than 10 times



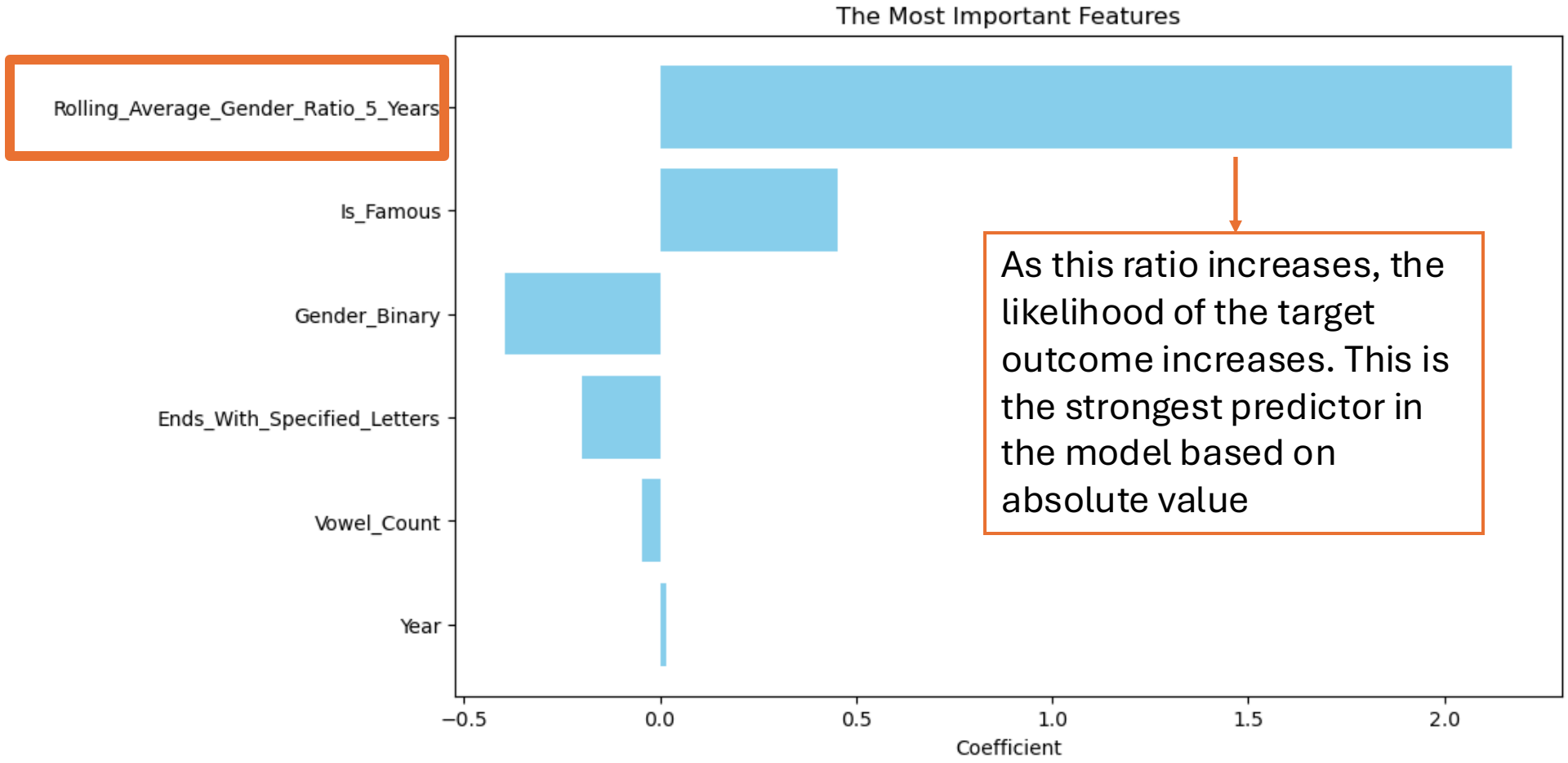
Top 1000 names vs. Popular Name Occurances More than 1000 times



Logistic Model

Model 1(99%)	All Features **(Overfit)**
Model 2 (96%)	Reduced Features, Mix of Binary and Scaled **(Overfit)**
Model 3 (96%)	Year Baseline, Mix of Binary and Scaled **(Overfit)**
Model 4 (89%)	Yearly_Change_ Name_Ratio, Mix of Binary and Scaled **(Underfit)**
Model 5 (98%)	Yearly_Change_Gender_Ratio/Rolling_Average_National_Ratio_5_Years **(Moderate)**
Model 6 (98%)	Rolling_Average_National_Ratio_5_Years **(Moderate)**

Logistic Regression Model Coefficients



Demo Platform

App Features:

Interactive Exploration: Users can explore historical name trends through an intuitive app interface.

Name Prediction: The app predicts future popular baby names based on historical data and current trends.

Demonstrate the app’s functionality, showcasing key features like name search, trend analysis, and prediction outputs.

Next Steps

1

Advanced Modeling: Improve predictive accuracy with advanced machine learning techniques and real-time data updates.

2

Time Series Model: Modify the dataset for time series modeling by incorporating time-dependent features, setting the time variable as the index, and creating lag features and rolling averages.

3

User Testing: Conduct user testing for the demo platform to refine the user experience and add new features.

THANKS FOR YOUR LISTENING!