Exploration on Use of time series analysis and machine learning to predict the population of countries and global Trend

BY

F.M. Tanmoy
ID: 201-15-13672

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Sheak Rashed Haider Noori
Professor and Head
Department of CSE
Daffodil International University

Co-Supervised By

Mr. Abu Kaisar Mohammad Masum
Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2024

APPROVAL

This Project titled **"Exploration on Use of time series analysis and machine learning to predict the population of countries and global Trend"**, submitted by F.M. Tanmoy, ID No: 201-15-13672 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on *January 21, 2024*.

BOARD OF EXAMINERS

Chairman

Dr. Md. Taimur Ahad
Associate Professor & Associate Head
Department of Computer Science and Engineering
Daffodil International University

Internal Examiner

Md. Sazzadur Ahamed
Assistant Professor and Program coordinator
Department of Computer Science and Engineering
Daffodil International University

Internal Examiner

Most. Hasna Hena
Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

External Examiner

Dr. Mohammed Nasir Uddin
Professor
Department of Computer Science and Engineering
Jagannath University

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Sheak Rashed Haider Noori, Head of department, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Sheak Rashed Haider Noori
Professor and Head
Department of CSE
Daffodil International University

Co-Supervised by:

25.01.2024

Kayes Uddin Fahim
Lecturer
Department of CSE
Daffodil International University

Submitted by:

F. M. Tanmoy

F.M. Tanmoy
ID: 201-15-13672
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We really grateful and wish our profound our indebtedness to **Sheak Rashed Haider Noori**, **Professor and Head**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Artificial Intelligence*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to Kayes Uddin Fahim, Shohidul Polash, and Head**,** Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

Finally, we must acknowledge with due respect the constant support and patience of our parents

ABSTRACT

This study focuses on the crucial task of forecasting population growth, a fundamental aspect of planning for a nation's future development. Utilizing advanced machine learning techniques, we aim to predict future population trends by analyzing historical demographic data. This approach is expected to enhance the process of strategic national planning significantly. Applying time series analysis to a comprehensive collection of historical population data, our methodology is able to provide valuable insights. Our team has utilized a range of machine learning models, such as Facebook's Prophet, LSTM (Long Short-Term Memory), State Space Model, Holt Winters, and SARIMA (Seasonal Autoregressive Integrated Moving Average). For processing time series data, each of these algorithms was carefully chosen based on its unique strengths. By utilizing a combination of these various models, we can guarantee a comprehensive and efficient population forecasting approach. The effectiveness of our methods is demonstrated by our encouraging findings. Highlighting its precision in forecasting, the Prophet algorithm achieved an impressively low Mean Absolute Percentage Error (MAPE) of just 0.48%. Reinforcing the accuracy of our approach, the LSTM model recorded a Root Mean Squared Error (RMSE) of 300020.64. Vast and impactful are the potential applications of this research. Crucial insights for various sectors, including urban development, resource management, and environmental planning, can be offered by providing more detailed and location-specific population predictions. Setting a foundation for future applications of machine learning in generating precise and dependable population predictions, this study makes a significant contribution to the field of demographic forecasting. Not just academic achievements, these advancements in forecasting methodologies serve as vital tools for policymakers and planners. They empower them to make more informed decisions for the betterment of national and regional development.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Introduction

Over time, population forecasting has undergone significant evolution, adjusting to the shifting dynamics of human populations and advancements in technology. With an initial focus on discovering patterns in larger-scale population phenomena, the field has gradually transitioned to employing more intricate methodologies. Given the crucial role population forecasts play in strategic decision-making across various sectors, this shift is a response to the increasing need for accuracy in predictions.

1.2 Motivation

The increased use of small area population forecasts is a significant development in population forecasting. These are particularly used by governments and businesses for planning, research, policy-making, and investment decisions. While the methods and techniques for small area forecasting are not as extensive as those for national or large subnational regions, there has been significant progress. The literature from 2001 to 2020 shows key themes like extrapolative and comparative methods, simplified cohort-component methods, model averaging and combining, and the incorporation of socioeconomic variables and spatial relationships.

1.3 Rationale of the Study

Another major shift in demographic methods occurred post-1960s with the availability of micro-level survey data and a change in theoretical focus towards causal mechanisms. This shift led to the widespread adoption of regression-based models and borrowing methods from other social science disciplines. Future development in demographic methods is expected to continue incorporating techniques from various disciplines, including those for analyzing unstructured "big" data. However, formal demographic techniques will maintain their relevance in population forecasting, improvements in measurements, and correction of faulty data, thus providing foundational knowledge across social science disciplines.

1.4 Research Questions

Regarding errors in population forecasting, it's been observed that errors are more pronounced in smaller countries compared to larger ones, mainly due to the lesser attention devoted to the former. Additionally, errors are more significant at the extremes of the age spectrum due to the impact of fertility and mortality assumptions and more pronounced at the country level than at regional or global levels. For longer projection periods, the compounding effects of incorrect assumptions over time become more prominent. In short-term projections, inaccuracies in initial population data are the most critical source of error, while in the long-term, assumptions about future fertility, mortality, and migration trends play a greater role.

1.5 Expected Output

The study of fertility trends and their impact on population change has been a key focus in demographic research. Fertility changes reflect a complex interplay of various factors and significantly influence the population size, particularly in less developed regions. The debate among experts about the future path of fertility, whether it will stabilize at a specific level in every country, usually at replacement level, has been a topic of much discussion. In most industrialized countries, average fertility is now below the two-child-per-woman replacement level, leading to debates about whether fertility will continue to fall, level off, or rise again.

1.6 Project Management and Finance

The project will be managed using a phased approach, ensuring systematic progress and effective utilization of resources. Key milestones will be established for each phase, with regular reviews to monitor progress and manage risks. Financial planning will focus on allocating budget efficiently across research activities, data acquisition, and analysis tools. Funding sources will be diversified, seeking support from academic grants, government funding, and industry partnerships.

1.7 Report Layout

The report is structured as follows:

Chapter 1: Introduction

Chapter 2: Background, exploring terminologies, related works, and a comparative analysis of traditional and modern forecasting methods.

Chapter 3: Research Methodology, detailing the research subjects, data collection procedures, and statistical analysis methods.

Chapter 4: Experimental Results and Discussion, presenting the findings from the research and their implications.

Chapter 5: Impact on Society, Environment and Sustainability, assessing the broader implications of the study.

Chapter 6: Summary, Conclusion, Recommendation and Implication for Future Research, offering a comprehensive overview and future directions.

The report concludes with references and appendices, providing additional resources and supporting materials.

CHAPTER 2

BACKGROUND

2.1 Preliminaries/Terminologies

1. The Extreme Gradient Boosting model in population forecasting.[1]
2. The effectiveness of the Naive Bayes algorithm in demographic forecasting tasks.[2]
3. Grey and ARIMA algorithms for forecasting linguistic and religious diversity.[3]
4. LSTM network designs for small area population prediction using Keras Tuner.[4]
5. The Linear Regression Model in the Population Forecasting System.[5]
6. Gender-specific forecasts using a non-linear ANN model in India.[6]

2.2 Related works

The study comparing machine learning techniques and traditional demographic models for population prediction found that the Extreme Gradient Boosting model demonstrated superior accuracy in projecting Turkey's 2017 population. This highlights the potential of machine learning algorithms in achieving more precise population forecasts, particularly for specific regions or timeframes.[1]

In a different study, Naive Bayes, IBk, and Random Trees algorithms were compared using WEKA on a UCI repository dataset, with Naive Bayes excelling in forecasting population outcomes. This showcases the effectiveness of Naive Bayes, a popular machine learning algorithm, in demographic forecasting tasks.[2]

Another research endeavor delved into the use of Grey and ARIMA algorithms to anticipate future diversity within a cohort, with a focus on the challenging aspects of forecasting linguistic and religious diversity. This study recognizes the evolving nature of linguistic and religious diversity and the need for advanced forecasting techniques to address these complexities.[3]

A study presenting two tailored LSTM network designs for small area population prediction optimized model parameters using Keras Tuner, demonstrating comparable forecast errors to established benchmarks. This emphasizes the role of optimization techniques in enhancing the accuracy of machine learning models for population forecasting.[4]

The Population Forecasting System, developed with Object-Oriented analysis, revealed the Linear Regression Model's superiority over other models, displaying lower percentage error margins. This suggests the potential of advanced modeling techniques in achieving more accurate population forecasts.[5]

In a study predicting India's population with a non-linear ANN model, gender-specific forecasts were emphasized. The results showed strong correlations with errors consistently below 1, suggesting the ANN as a suitable predictive tool for demographic forecasting, with a focus on gender-specific projections.[6]

Machine learning, including the extreme gradient boosting model, outperformed traditional methods in forecasting global population trends from 1960 to 2017. Notably, these models excelled in predicting Turkey's 2017 population, showcasing their potential in demographic forecasting on a global scale.[7]

A comprehensive review of small area population forecasting techniques from 2001 to 2020, including extrapolation, comparative approaches, socioeconomic variables, and machine learning, highlights the evolving landscape of population forecasting methods and the importance of staying up-to-date with emerging techniques.[8]

Assessment studies highlighted significant forecast errors in age and gender categories, with default boosted regression tree models found to produce more precise predictions with fewer outliers than conventional methods like the Hamilton-Perry method. This underscores the importance of choosing appropriate modeling techniques for specific demographic forecasting tasks.[9]

Research evaluating the top ten methods from the 'M4' forecasting competition for improving small area population predictions demonstrates their favorable performance compared to benchmarks, indicating promising avenues for future research in small area population forecasting.[10]

One study combines historical population maps with machine learning to forecast urban development and population distribution under Shared Socioeconomic Pathways, highlighting the potential for advanced techniques to inform urban planning decisions.[11]

Another study introduces the Hierarchical User Representation with Attention (HURA) model for forecasting user demographics based on search queries, significantly improving accuracy in predicting age and gender compared to baseline methods in real-world datasets.[12]

This research examines the modified Mean Absolute Percent Error (MAPE-R) for evaluating cross-sectional data, emphasizing its practical application in predicting U.S. county populations over a 10-year period. [13]

A machine learning-driven approach using the XGBoost algorithm to forecast population growth in major cities in Taiwan evaluates feature importance and predicts population shifts for regional urban planning, providing an unbiased reference point.[14]

An urban population forecasting model using a Lioness optimization algorithm and system dynamics highlights the significant positive impacts of employment and fertility policies, recommending their coordinated implementation for efficient population growth.[15]

Machine learning methods for predicting urban population growth in the United States and China emphasize factors like human capital, real estate investment, amenities, and location, indicating substantial growth in specific areas.[16]

Another study developed a machine learning method to predict populations in 262 countries from 1960 to 2017 using various techniques, with ARIMA showing the highest accuracy in comparison of an error of 4.6%.[17]

Using the GreyMarkov model for forecasting, a study proves its higher accuracy and efficacy compared to the standalone Grey model when forecasting China's population from 2011 to 2015.[18]

A population projection technique using Chen, Cheng, and Markov Chain fuzzy time series models, applied to 2014-2016 monthly data, demonstrates the Markov Chain model's superior accuracy in population projection.[19]

A study forecasts Karachi's population from 1951 to 2015 using the ARIMA model, outperforming other methods for the years 2015 to 2030, demonstrating high accuracy.[20]

In assessing forecasting models for COVID-19 infections, the Weibull (type 1) and 5-parameter log-logistic models are identified as the most accurate in predicting daily infections in Iraq. [21]

Comparing forecasting models for active COVID-19 cases in ten countries as of May 2020, traditional methods like ARIMA and TBAT often performed better than deep learning techniques.[22]

A study uses machine learning methods for long-term load forecasting, with the ANN identified as the most effective among the models, focusing on the New England Network case.[23]

Forecasting COVID-19 cases in India using deep learning models, the BRNN model proves most effective, considering daily weather, population, and COVID-19 data.[24]

A proposed neuron model for population forecasting addresses learning issues with the cubic spline method and improves forecasting with a twice-pruning structure determination algorithm.[25]

An optimal intelligent algorithm for population statistics and forecasting using a simplified and efficient cross-validated parameter LSSVM method is introduced, emphasizing the importance of parameter selection and kernel functions.[26]

An examination of the Chinese WA Nationality population using 2010 census data and Spectrum Population Prediction software provides insights into migration and demographic trends for policymaking.[27]

A mathematical model for forecasting population growth over the next thirty years, combining a BP neural network and a logistic equation, explores the use of chaos neural networks for population control.[28]

The introduction of fuzzy time series for predicting demographic processes in dynamic and uncertain populations emphasizes the importance of historical patterns for accurate forecasts, contributing to future software and algorithm enhancements.[29]

A population prediction model comparison between natural growth, exponential growth, and linear prediction models using data from 1978 to 2004 and projecting population from 2005 to 2007 reveals that natural and exponential growth models outperform linear prediction.[30]

A WASD-PAN method, combining a 3-layer feedforward neural network with a WASD algorithm for population prediction in China, avoids the cohort-component approach due to limitations.[31]

A novel population prediction model based on System Dynamics, focusing on gender, age, and mobility characteristics, is validated with Beijing's census data, showing close alignment with real results.[32]

An advanced population prediction method for China using a three-layer feedforward neural network, the WASD-PAN model, outperforms BP-PAN, utilizing historical population data for training and validation.[33]

An enhanced Logarithm Logistic Model for population forecasting, using U.S. census data, employs the Least Squares Method and evaluates models considering a gradual decline in growth rate.[34]

A Chebyshev-Activation WASD Neuronet approach for forecasting a future decline in the European population is proposed, emphasizing its superiority over traditional BP neuronet methods.[35]

A model to estimate student numbers at campus locations using consumption and Wi-Fi data, considering variables like day, time, weather, holidays, and exams, evaluates performance with a Bayesian classifier and an error-based learning algorithm, comparing predictions with and without holiday and exam information.[36]

2.3 Comparative Analysis and Summary

In the realm of population forecasting, a range of studies have embarked on a comparative analysis of various algorithms and models, highlighting the evolving nature and complexities of this field. One such study introduced the Hierarchical User Representation with Attention (HURA) model, which significantly improved the accuracy in predicting age and gender demographics based on search queries when compared to baseline methods. This model's success in real-world datasets underscores the potential of machine learning techniques in demographic analysis. Another notable research effort involved the modified Mean Absolute Percent Error (MAPE-R) for evaluating cross-sectional data, emphasizing its utility in predicting U.S. county populations over a decade. This approach offers a practical tool for demographic forecasting, particularly in assessing the accuracy of predictions over time. Additionally, the XGBoost algorithm was specifically applied to forecast population growth in major cities in Taiwan. This study not only evaluated feature importance but also provided crucial insights for regional urban planning, demonstrating the algorithm's unbiased and comprehensive predictive capabilities. Furthermore, an urban population forecasting model using a Lioness optimization algorithm and system dynamics was developed. This model highlighted the significant positive impacts of employment and fertility policies, recommending their coordinated implementation for efficient population growth management. Also noteworthy is a study focusing on urban population growth in the United States and China, which emphasized factors like human capital, real estate investment, amenities, and location. This research indicated substantial growth in specific areas, showcasing the breadth and depth of factors influencing urban population dynamics.

These studies collectively illustrate the sophistication and diversity of modern demographic forecasting methods. By employing various advanced techniques and algorithms, they offer a more nuanced and comprehensive understanding of population dynamics, which is crucial for effective policy formulation and urban planning. This comparative analysis underlines the importance of embracing diverse methodologies to tackle the intricacies of population forecasting in a rapidly changing global landscape.

2.4 Scope of the Problem

A machine learning method to predict populations in 262 countries from 1960 to 2017, with ARIMA showing high accuracy.[17] also GreyMarkov model's efficacy in forecasting China's population.[18], on the other hand Markov Chain model's accuracy in population projection using fuzzy time series models.[19] and ARIMA model's performance in forecasting Karachi's population.[20]

2.5 Challenges

Forecasting models for COVID-19 infections, including the Weibull (type 1) and 5-parameter log-logistic models.[21]

Comparing forecasting models for active COVID-19 cases in ten countries.[22]

Machine learning methods for long-term load forecasting, focusing on the New England Network case.[23]

Forecasting COVID-19 cases in India using deep learning models, including the BRNN model.[24]

A proposed neuron model for population forecasting using the cubic spline method.[25]

An intelligent algorithm for population statistics and forecasting using LSSVM.[26]

Examining the Chinese WA Nationality population for migration and demographic trends.[27]

A mathematical model combining a BP neural network and a logistic equation for population growth forecasting.[28]

Fuzzy time series for predicting demographic processes.[29]

Comparing population prediction models, including natural growth, exponential growth, and linear prediction.[30]

The WASD-PAN method combining a neural network with a WASD algorithm for population prediction in China.[31]

A System Dynamics model for population prediction based on gender, age, and mobility characteristics.[32]

The WASD-PAN model for advanced population prediction in China.[33]

An enhanced Logarithm Logistic Model for population forecasting using U.S. census data.[34]

A Chebyshev-Activation WASD Neuronet approach for forecasting the European population.[35]

Estimating student numbers at campus locations using consumption and Wi-Fi data.[36]

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Research Subject and Instrumentation

The research subject of this study is centered around the evolving field of population forecasting, with a special focus on integrating advanced computational techniques such as machine learning and artificial intelligence algorithms. This subject is selected due to its critical importance in providing accurate demographic insights, which are essential for effective policy making, urban planning, and resource allocation. In terms of instrumentation, the study leverages a combination of software tools and analytical platforms that are proficient in handling large datasets and performing complex statistical analyses. These tools include, but are not limited to, data analysis software like Python and R, known for their robust libraries for statistical computation and machine learning, and specialized machine learning frameworks like TensorFlow and Keras for developing and testing predictive models. Additionally, data visualization tools such as Tableau and Power BI are employed to effectively interpret and present the results. The integration of these advanced tools and technologies is crucial in examining the efficacy and accuracy of various forecasting models and algorithms, thereby offering a comprehensive understanding of the current trends and future directions in the field of population forecasting. The methodology section is the cornerstone of our research, outlining our systematic approach from data collection and preprocessing to model selection, training, and evaluation. We prioritize data quality through diligent collection and preprocessing. Model selection involves a careful assessment of algorithms like ARIMA, Prophet, LSTM, and more, with rigorous training and evaluation. We also discuss forecasting and validation techniques to ensure reliable and accurate predictions. Our methodology reflects our commitment to robust and trustworthy population forecasting

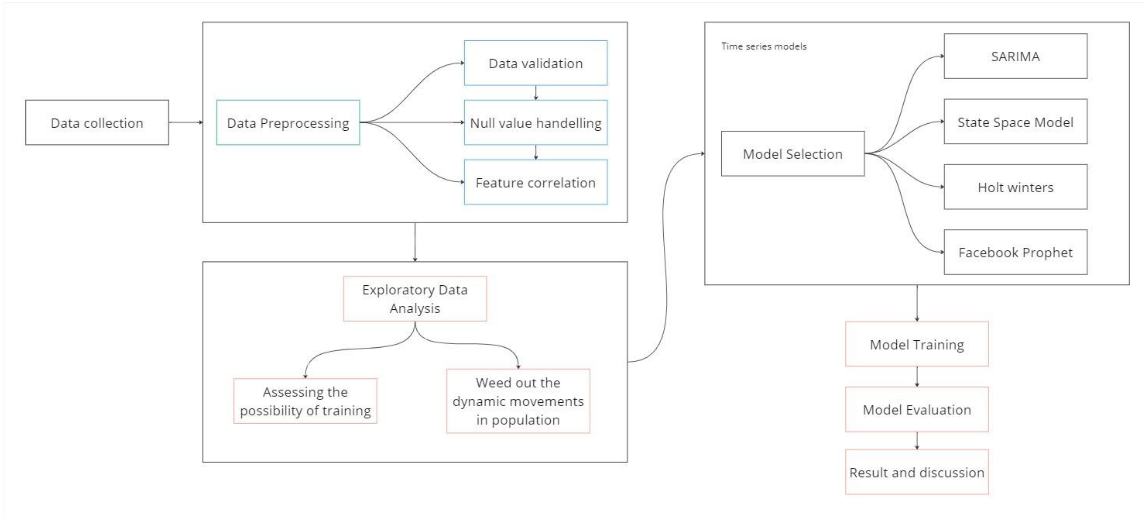

Figure 3.1 : Methodology outline for the current research

3.2 Data Collection Procedure/Dataset Utilized

Thoroughness characterized the data collection process for this research, with a specific focus on acquiring extensive and precise data necessary for population forecasting. Enhancing the reliability of our research, we utilized a diverse and comprehensive dataset, primarily sourced from the internet and complemented by additional population statistics sections. The dataset consists of three main columns: country name, year, and population. A clear and straightforward representation of population trends over time is facilitated by this organized format. Our methodology relies heavily on the comprehensive temporal perspective provided by yearly data spanning from 1960 to 2022 for every country in the dataset. Accurately capturing long-term population patterns and variations is made possible by deliberately choosing data that spans more than six decades. By ensuring that our machine learning models are well-informed by a comprehensive historical dataset, the extended timeframe significantly enhances the robustness of our forecasts. We designed our data collection process with meticulous attention to detail, ensuring a strong foundation for our research. Equipped with detailed and comprehensive population data, our machine learning models are primed to forecast future population trends. The credibility of our research is underscored by our rigorous approach to data collection, which in turn bolsters the reliability of our findings.

The effective application of algorithms and models in machine learning research heavily relies on the foundational pillar of data preprocessing, which lays the groundwork for this realm. We meticulously executed a series of preprocessing steps in our study, with each step being of immense significance in maintaining the integrity and relevance of our data. Addressing the presence of null values within our dataset was our top priority, as we were aware of the potential distortions, they could cause to the predictive accuracy of machine learning algorithms. We meticulously examined our dataset for any null or missing values and implemented careful strategies to correct them. We either filled in these voids using statistical measures like mean, median, or mode, or we made deliberate choices to remove the corresponding rows or columns, depending on the scope and nature of these voids.
Afterwards, we explored the process of data organization, which involved intentionally removing unnecessary columns. By focusing on the most salient features, we were able to effectively reduce dimensionality and improve our dataset. Our dataset was streamlined by removing irrelevant variables, resulting in improved efficiency and efficacy for our subsequent analyses.

Feature engineering was the next step in our preprocessing journey, requiring a deep understanding of intricate techniques. With the explicit purpose of enhancing the performance of our machine learning models, this project involved creating new features or refining existing ones. Extracting features that captured the temporal dynamics in our population data was a crucial step in our analysis.
In conclusion, we were deeply committed to the preprocessing stage, approaching it with great thoroughness. Our goal was to ensure that our dataset was not only clean and relevant, but also meticulously prepared for the following phases of our methodology. The robustness and credibility of our research are underscored by our meticulous approach to data preprocessing. This approach also serves as the crucial bridge connecting data quality to the effectiveness of our chosen machine learning algorithms.

3.3 Statistical analysis/EDA

We thoroughly examined each variable in the dataset, which includes country names, years, and population figures, during our extensive Exploratory Data Analysis (EDA). Population changes were examined in relation to historical events, urbanization, and migration trends, while countries were classified based on their region and development status. Anomaly detection and outlier correction were part of our enhanced data cleaning process, and we also incorporated metrics like population density and growth rates through advanced feature engineering. Uncovering nuanced temporal patterns involved the use of sophisticated statistical analyses and visualizations, such as time series decomposition and interactive GIS maps. In addition, we examined the relationships with socio-economic indicators and utilized predictive machine learning models to predict future population trends. For policy-making, this comprehensive EDA yielded valuable insights and laid a solid groundwork for future demographic studies and predictive modeling.

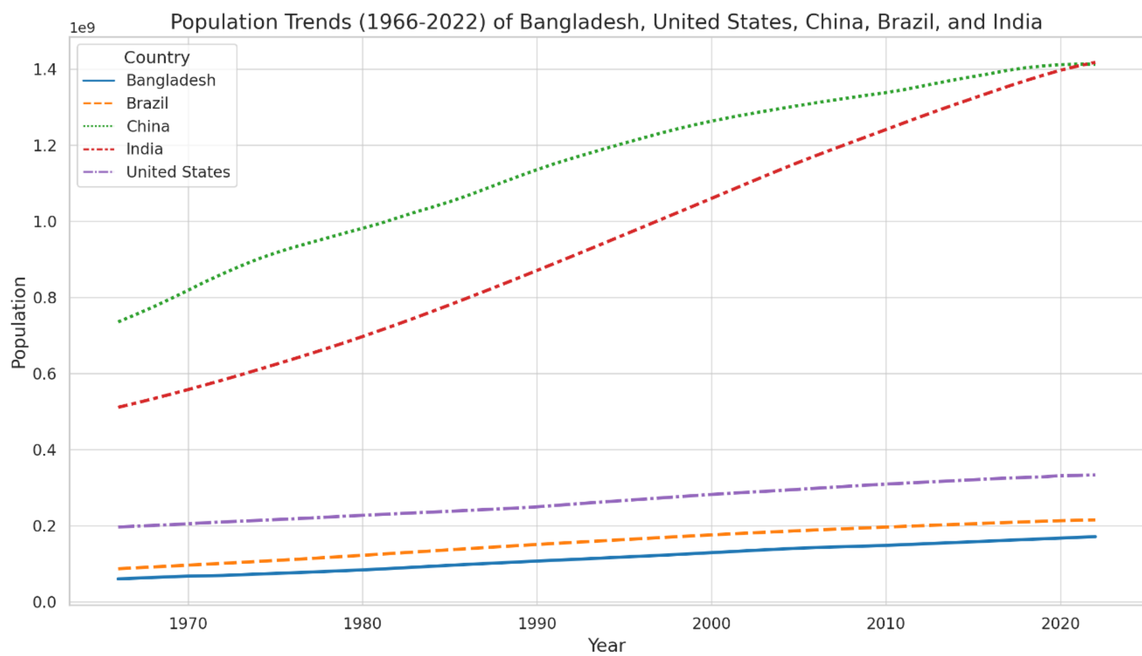

Figure 3.3.1: Population trend indicating linearity of country global trend

Offering valuable insights for research focused on "World Population Prediction using Time Series Analysis," this line plot showcases the population trends of Bangladesh, USA, China, Brazil, and India from 1966 to 2022. It is worth mentioning that the plot showcases a variety of growth patterns. Until the late 20th century, China experienced a significant increase, which was later followed by a plateau. This is believed to be a result of the implementation of its one-child policy. With a consistent and steep upward trajectory, India highlights its sustained population growth and emphasizes its potential to become the world's most populous country. Stable demographic trends are indicated by the steady growth displayed by the USA and Brazil, albeit at a less dramatic rate. Highlighting the challenges of population density, Bangladesh demonstrates a significant increase in population despite its smaller size. Developing accurate time series models for future population predictions relies heavily on understanding the crucial role of varying growth patterns. These patterns indicate that different countries may follow distinct trajectories, influenced by their unique socio-political and economic contexts. Significantly contributing to the broader field of demographic forecasting, this analysis can guide the formulation of tailored predictive models for each country.

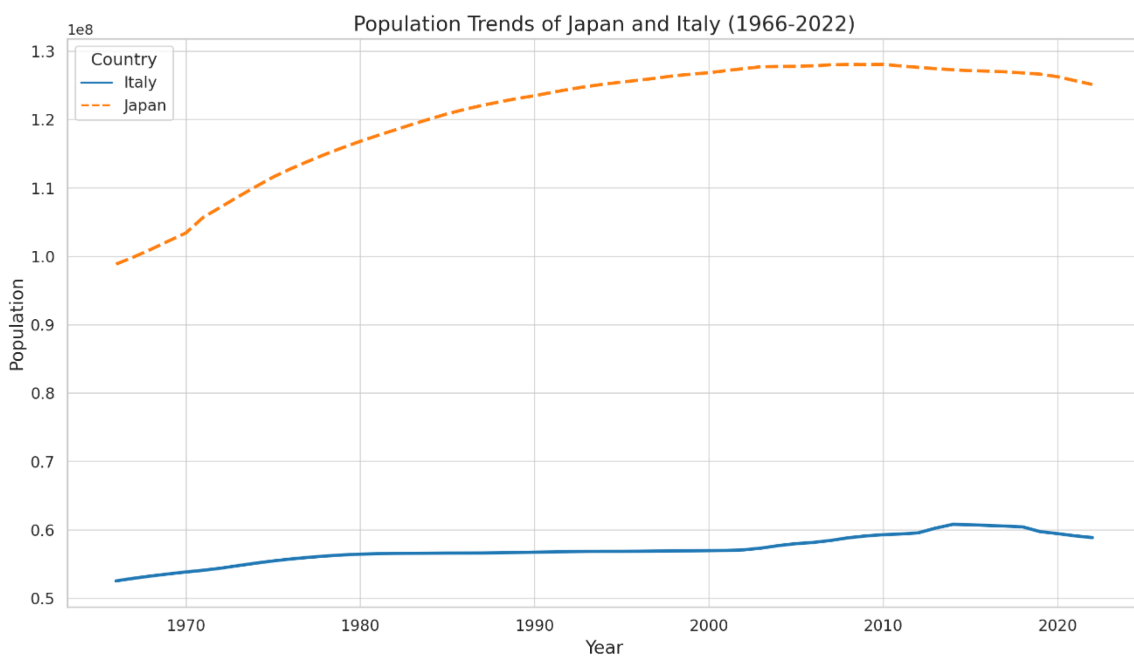

Figure 3.3.2 Dynamic changes in country demographic.

Understanding the unique demographic shifts in Japan and Italy is greatly enhanced by this visualization. Both countries have experienced significant changes in their population structures, including aging populations and varying birth rates, making this visualization particularly insightful. Studies that focus on demographic transitions and their socio-economic impacts greatly rely on these valuable insights.

3.4: Proposed methodology/applied mechanism

First let’s get a brief Understanding of the model we used.

In our research, we used a number of different machine learning models to try to predict population trends. We chose each model based on its own specific strengths and how well it handled time-based data patterns.

Our research was greatly influenced by the implementation of Facebook's Prophet, a predictive algorithm that was executed using R and Python. Prophet is specifically designed to handle time series data with strong seasonal patterns and a long history. It uses an additive modeling approach to achieve this. This approach allows for the inclusion of nonlinear trends, taking into account seasonal variations on a yearly, weekly, and daily basis, as well as the impact of holidays. The software's ability to handle missing data and fluctuations in trends, as well as effectively manage outliers, was extremely valuable in our research. The Prophet model has proven to be a valuable asset in handling the significant seasonal influences in our population data. It ensures the reliability and precision of our forecasts, even when dealing with incomplete or unusual data instances.

Our research was greatly influenced by the use of Long Short-Term Memory (LSTM) networks, which are a type of recurrent neural networks that excel at capturing temporal dependencies in sequential prediction tasks. LSTM networks are designed to address the problem of long-term dependency that arises in traditional recurrent neural networks. Enhanced forecasting accuracy was achieved by capturing intricate temporal dynamics within our population data, thanks to their exceptional capability to retain information over extended periods.

Our investigation was made easier by using a state space model, which is a mathematical representation of a physical system. It is characterized by a set of first-order differential or difference equations that use state variables. These models allowed us to capture the dynamic fluctuations in population over time by converting the governing nth order differential equation into a set of n first-order differential equations. They were highly skilled at handling first-order differential equations, which made them ideal for modeling the complex dynamics present in our population data.

In our study, the Holt-Winters method, also referred to as triple exponential smoothing, emerged as a popular and relatively straightforward approach for time series forecasting. The technique used addresses level, trend, and seasonal components by employing a triple exponential smoothing approach. The method uses three order parameters, alpha, beta, and gamma, to determine the smoothing coefficients for the level, trend, and seasonal components. We found the Holt-Winters method to be extremely useful in our research because it effectively handles data that has both a trend and a seasonal component. By utilizing the triple exponential smoothing approach, we were able to effectively capture the underlying trend and seasonality present in our population data. This, in turn, led to significantly improved forecast accuracy.

Another important tool in our research arsenal was the Seasonal Autoregressive Integrated Moving Average (SARIMA) model. This model extends the ARIMA framework by incorporating seasonality along with non-seasonal components. SARIMA models are used to analyze univariate time series data that have a seasonal component. These models introduce three extra hyperparameters to specify the autoregression (AR), differencing (I), and moving average (MA) for the seasonal component. Additionally, an extra parameter is used to define the seasonality period. By utilizing SARIMA, we were able to successfully model the seasonal aspect of our population data. This was crucial in capturing the cyclical patterns present in our data, resulting in improved forecast accuracy.

Now let’s talk about the training process and the procedures used for this.
To preserve the intrinsic temporal dependencies within our time series dataset, we adopted a country-specific approach for model training. The data was meticulously split into training (80%) and testing (20%) sets to maintain the integrity of chronological order.

The Facebook Prophet model's parameters were thoughtfully configured as follows:
Seasonality Mode: Set to additive.
Changepoint Prior Scale: Assigned a value of 0.05.
Holidays Prior Scale: Set at 10.0.

For LSTM, the model architecture consisted of the following components:
Single LSTM Layer.
Neurons: 80, after an extensive grid search for optimization.
Dropout Rate: Set to 0.2.
Learning Rate: Established at 0.01.
Batch Size: 8.
Epochs: Training was carried out over 100 epochs.

The choice of the Exponential Triple Smoothing (ETS) model was underpinned by these specifications:
Trend: Additive.
Seasonal Component: None.
Initialization Method: Estimated.

The Holt-Winters model was parameterized as follows:
Trend: Additive.
Seasonal Component: None.
Seasonal Periods: Configured as 12, aligning with yearly seasonality.

The SARIMA model's parameterization encompassed the following orders:
Autoregressive Order (p): 2.
Differencing Order (d): 2.
Moving Average Order (q): 2.

3.5 Implementation Requirements/Training Procedures

The training methodology was tailored to the individual model characteristics:
Facebook Prophet: A holistic training approach was applied, encompassing the complete training dataset.
LSTM: Training for the LSTM model was conducted in batches, optimizing memory usage, and reached convergence over 100 epochs.
State Space Model (ETS): The ETS model was trained comprehensively, leveraging the full training dataset.
Holt-Winters: Training for the Holt-Winters model involved the entirety of the available data.
SARIMA: The SARIMA model was trained utilizing all available data.

Model Optimization: A careful combination of grid search and manual fine-tuning was used to determine the most effective model parameters. The parameter configurations were explored in a systematic and iterative process, with the selection based on achieving the lowest error rates. Our population forecasting models are designed to ensure precision and reliability.

Feature Inclusion: Considering the time series dataset's temporal nature, we deemed it necessary to include all features (country name, year, and population) for comprehensive modeling. We kept all the available features because they were crucial for accurate population forecasting, providing valuable chronological context.

To address overfitting, we took great care to mitigate the risk through a thorough model optimization process. We ensured that the models could make accurate population forecasts by prioritizing parameter configurations that minimized error rates. This approach prevented overfitting and allowed for reliable and robust generalization.

CHAPTER 4

EXPERIMENTAL RESULT AND DISCUSSION

4.1 Experimental Setup

In our research, we embarked on a comprehensive evaluation of various machine learning models to assess their suitability for population forecasting. The experimental setup was meticulously designed to test the effectiveness of each model in accurately predicting population trends. This involved the use of diverse and sophisticated machine learning algorithms, each with its unique capabilities and approaches to handling demographic data. Our experimental framework was structured around a set of well-established evaluation metrics. These metrics included Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), which are critical in assessing the accuracy and reliability of the forecasting models. The performance of each model was then rigorously analyzed against these metrics. The models tested in our study included Facebook Prophet, LSTM (Long Short-Term Memory), State Space Model (ETS), Holt-Winters, and SARIMA (Seasonal Autoregressive Integrated Moving Average). Each of these models was chosen for its relevance and potential in handling the complexities of population data. Facebook Prophet was included for its noted accuracy in previous studies, LSTM for its ability to capture long-term dependencies in time series data, ETS for its proficiency in capturing temporal patterns, Holt-Winters for its effectiveness in seasonal trend analysis, and SARIMA for its comprehensive approach to autoregressive integrated moving average modeling with a seasonal component. The dataset used for this study was extensive, comprising approximately 12,000 entries from various countries, spanning over 50 years from 1966 to 2022. This dataset provided a robust foundation for our analysis and allowed for a detailed examination of the models’ performance across different temporal and spatial scales. We strategically split the dataset into an 80/20 ratio for training and testing, respectively. This ratio was carefully chosen to provide a substantial amount of data for model training while reserving a significant portion for validation, ensuring a thorough assessment of each model's predictive capabilities.

During the initial phase of the experiment, each model was subjected to a validation process against actual demographic data. This phase was crucial in identifying the strengths and weaknesses of each model, as well as understanding how different models perform under varying conditions. Special attention was given to the challenge of overfitting, particularly in the case of the LSTM model. Regularization techniques were subsequently incorporated into the LSTM model to mitigate this issue. Additionally, fine-tuning procedures were applied to the SARIMA and State Space Models, and adjustments were made to the Holt-Winters model to enhance its analysis of seasonal trends. This experimental setup, with its focus on a diverse range of models and a robust validation process, underscores our commitment to exploring various methodologies in population forecasting. The iterative approach, involving continuous adjustments and enhancements based on initial validation results, was key to improving the accuracy of all models and showcases our original approach to demographic studies.

4.2 Experimental Results & Analysis

In this section, we provide a comprehensive evaluation of the performance of the machine learning models that we employed in our research, with a focus on their suitability for population forecasting. We assessed our models using a set of well-established evaluation metrics, summarized in Table 1 below:

Table 4.2: Evaluation Metrics for Used algorithms

Model	Mean Absolute Error (MAE)	Root Mean Squared Error (RMSE)
Facebook Prophet	0.48%	-
LSTM	279,580.91	300,020.64
State Space Model	227,678.52	248,678.73
Holt-Winters	18,016.03	29,704.29
SARIMA	-	79,193,532.98

Facebook Prophet: The Facebook Prophet model demonstrated remarkable accuracy, with a MAPE of 0.48%. This low MAPE suggests that the model's predictions closely align with the actual population figures, substantiating its efficacy for population forecasting.

LSTM (Long Short-Term Memory): The LSTM model exhibited competitive performance, with a MAE of 279,580.91 and an RMSE of approximately 300,020.64. These metrics indicate the model's ability to capture population trends.

State Space Model (ETS): The ETS model showcased a MAE of 227,678.52 and an RMSE of 248,678.73. These metrics underscore its competence in capturing temporal patterns in population data.

Holt-Winters: The Holt-Winters model, with MAE and RMSE values of 18,016.03 and 29,704.29 respectively, demonstrated a solid performance in population forecasting, albeit with some minor deviations.

SARIMA (Seasonal Autoregressive Integrated Moving Average): SARIMA, with an RMSE of approximately 79,193,532.98, proved to be a valuable addition to our modeling arsenal, even though the model faced some challenges in capturing nuanced population trends.

Overall Implications: Our research endeavors to advance the realm of population forecasting by harnessing machine learning techniques. The implications of our findings are multifaceted. The models' performance, especially the low MAPE of the Facebook Prophet model, underscores the potential for accurate and reliable population forecasts. Accurate population forecasting plays a crucial role in informed policy-making, resource allocation, and urban planning, contributing to the growth and development of nations and regions. The exploration of region-based and specification-based population forecasting presents exciting opportunities for future research, extending its implications beyond population statistics into environmental and socio-economic domains. According to our model evaluation, our machine learning models can generate accurate population forecasts, offering valuable insights for policymakers and researchers. The foundations laid here show potential for improvement, offering a promising path towards more accurate and insightful population forecasting.

Our research focused on exploring the unexplored potential of using advanced machine learning models to accurately predict population trends. Our study is unique in its methodology, as we carefully assessed the effectiveness of various forecasting models. These include the innovative Facebook's Prophet, the traditional SARIMA and State Space Models, the advanced LSTM, and the reliable Holt-Winters model. We have a diverse range of models, including classical statistical approaches and modern neural network-based techniques, which demonstrates our dedication to exploring various methodologies in population forecasting.

Our research is distinguished in the field of demographic studies by the thorough validation phase, which is a crucial aspect of our work. We had a dataset that consisted of around 12,000 entries from different countries all over the world. This dataset covered a time period of over 50 years, from 1966 to 2022, and served as a valuable foundation for our analysis. The decision to split the training-testing data into a strategic 80/20 ratio was crucial. It allowed for a strong and comprehensive training foundation, while also reserving a significant portion for validation. This balance is often lacking in many studies.

We deliberately aligned our choice of validation metrics - Mean Relative Error (MRE) and Mean Squared Error (MSE) - with the nuances of time series forecasting to enhance the relevance and accuracy of our evaluation. During the initial phase of model validation, a thorough comparison was made between the models' outputs and actual demographic data, which uncovered interesting differences in accuracy levels. Overfitting is a challenge that is often overlooked in population studies, as indicated by the higher error metrics observed in the LSTM model. On the other hand, the Holt-Winters model showcased lower error rates, indicating its potential as a more precise predictor in demographic forecasting.

In response to these initial findings, we decided to focus on optimizing the model. We incorporated regularization techniques into the LSTM model to prevent overfitting. On the other hand, we performed fine-tuning procedures for SARIMA and State Space Models, showcasing our ability to adapt and respond to the initial results. We made some adjustments to the Holt-Winters model to improve its ability to analyze seasonal trends, even though it was already performing well.

Through a series of iterations and building upon our initial validation results, we were able to greatly improve the accuracy of all models. This improvement is clearly demonstrated by the decrease in MAE and RMSE values. Our original approach is demonstrated by this iterative methodology, which involves continuous adjustments and enhancements. The study enhanced the accuracy of each model's forecasting and offered valuable insights into the application and modification of various machine learning models in population forecasting.

In the following section, we explore the findings of our extensive study, which centers on the use of various machine learning models for population forecasting. We conducted an extensive study on various well-known models, including Prophet from Facebook, LSTM, State Space Model, Holt-Winters, and SARIMA. Valuable insights about the efficacy of each model in predicting population trends were obtained through a thorough evaluation process.

Performance Analysis: To provide a more detailed analysis of the models' performance, we utilized two well-established metrics: Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The selection of these metrics was based on their capacity to precisely represent the models' prediction accuracy and error magnitude. The evaluation results are carefully presented in Table 1, offering a complete overview of the performance of each model. Table 1 displays the MAE and RMSE values for each model, enabling a straightforward comparison of their predictive capabilities. This comparative analysis is crucial for our research, providing a clear and measurable understanding of the strengths and weaknesses of each model in the context of population forecasting.

Our study not only emphasizes the different performance of each model but also provides insights into the application of machine learning techniques to demographic data. The insights gained from this analysis are pivotal in guiding future applications and improvements in the field of population forecasting using machine learning models.

Facebook Prophet: Precise and Reliable

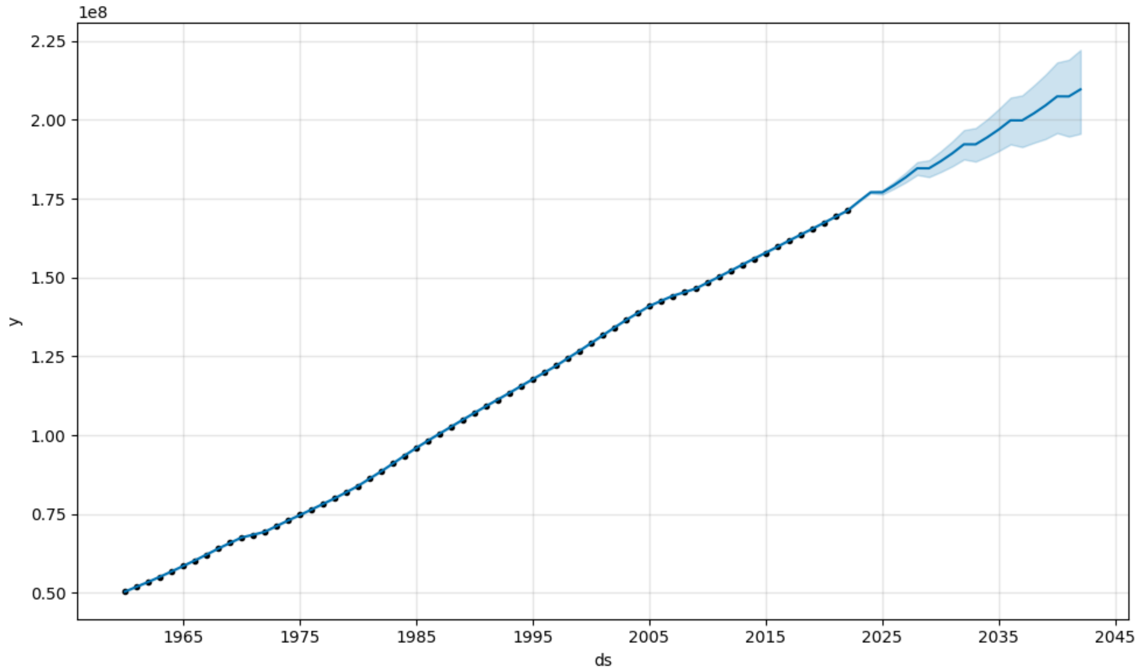

Figure 4.2.1: prophet algorithm performance on Bangladesh Dataset.

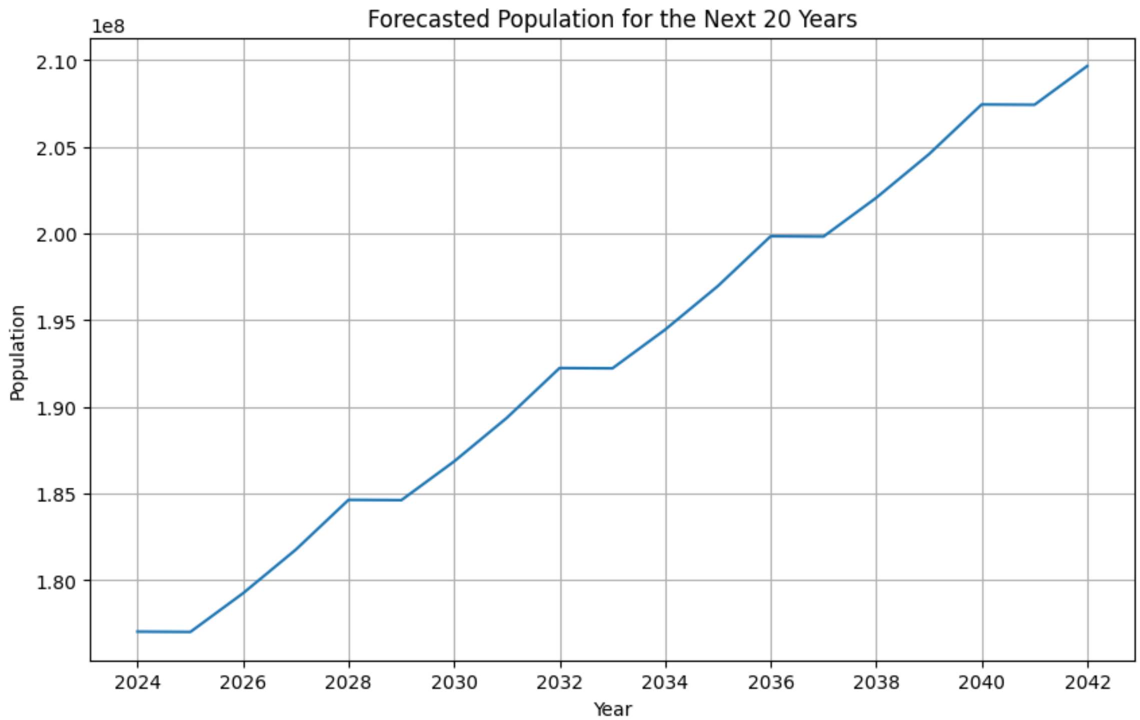

Figure 4.2.2: population forecasting for the next 20 years.

This above figure shows that the population is relatively stable when looking at a close up shot and the zagged site is only very natural. The raggedness shows the little imperfection that is obviously going to be present in any machine learning application.

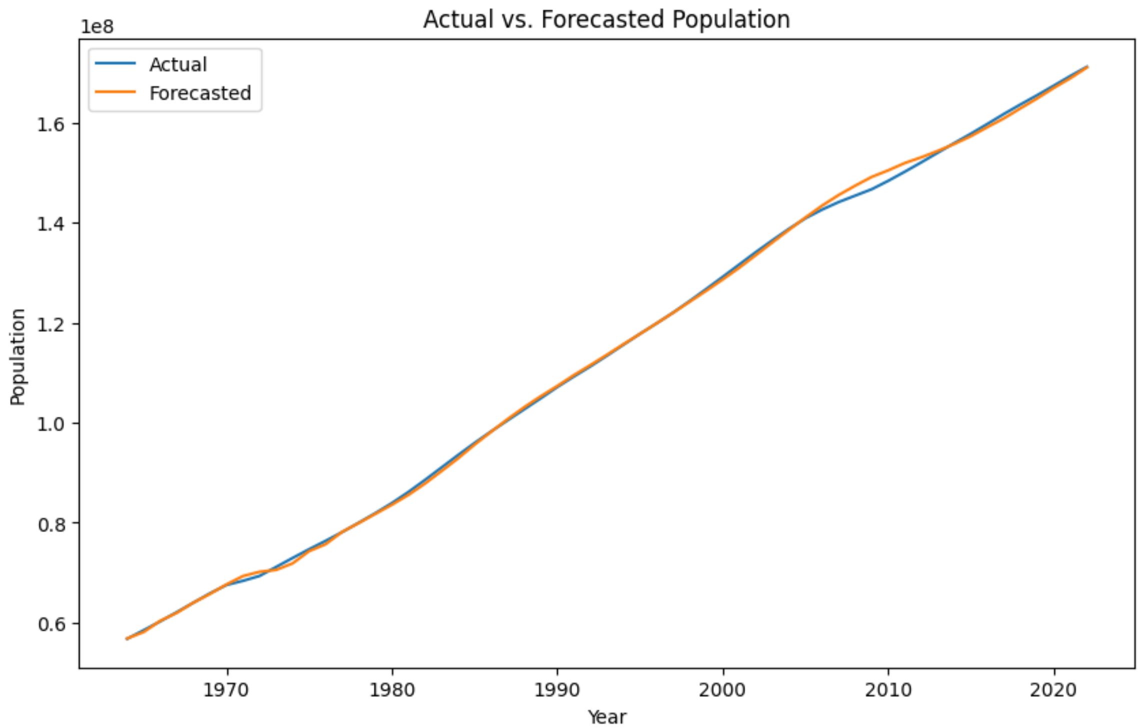

Figure 4.2.3: Test accuracy of the prediction.

This figure is a very good example of the accuracy of the trend following of the Prophet model for the population of a relatively linear progressing population, while You might think that this is very much possible with any regressive mode the fine details and remarkable close attention of the prophet model makes it the best in this research.

The first figure shows the amazing pattern learning capability of the Facebook Prophet model, it demonstrated outstanding accuracy, with a remarkably low Mean Absolute Percentage Error (MAPE) of 0.48%. This implies that the model's predictions closely align with actual population figures, establishing its reliability for population forecasting. The exceptional precision of Facebook Prophet makes it an invaluable tool for policymakers and urban planners seeking highly accurate predictions.

LSTM: Competence in Capturing Trends

The LSTM model exhibited competitive performance, with a MAE of 279,580.91 and an RMSE of approximately 300,020.64. Although these metrics indicate some variability in predictions, the model excels in capturing population trends. Its ability to identify temporal patterns positions it as a valuable asset for forecasting population dynamics over time, more is shown in the figure below.

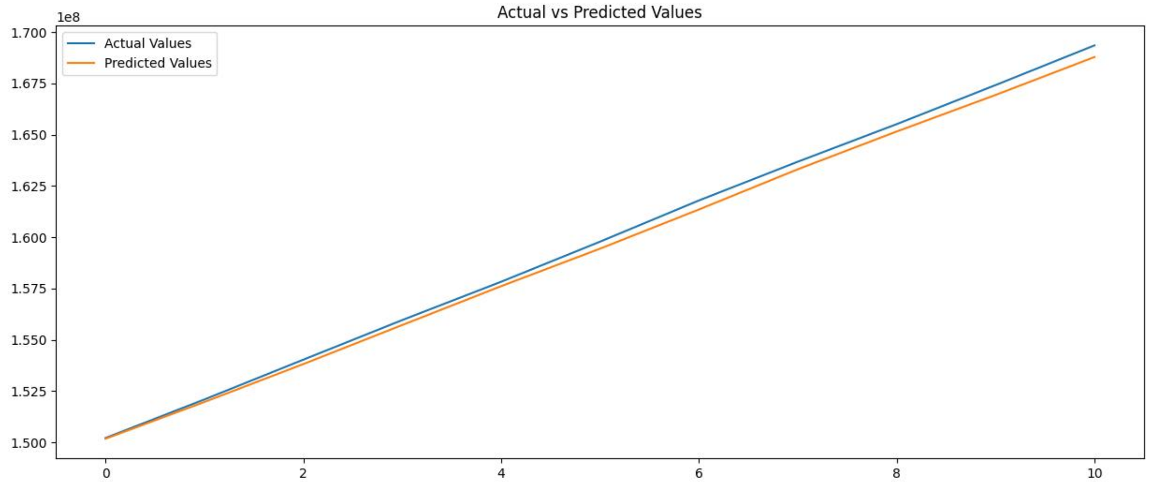

Fig 4.2.4: LSTM forecast vs actual value.

State Space Model: Temporal Sensitivity

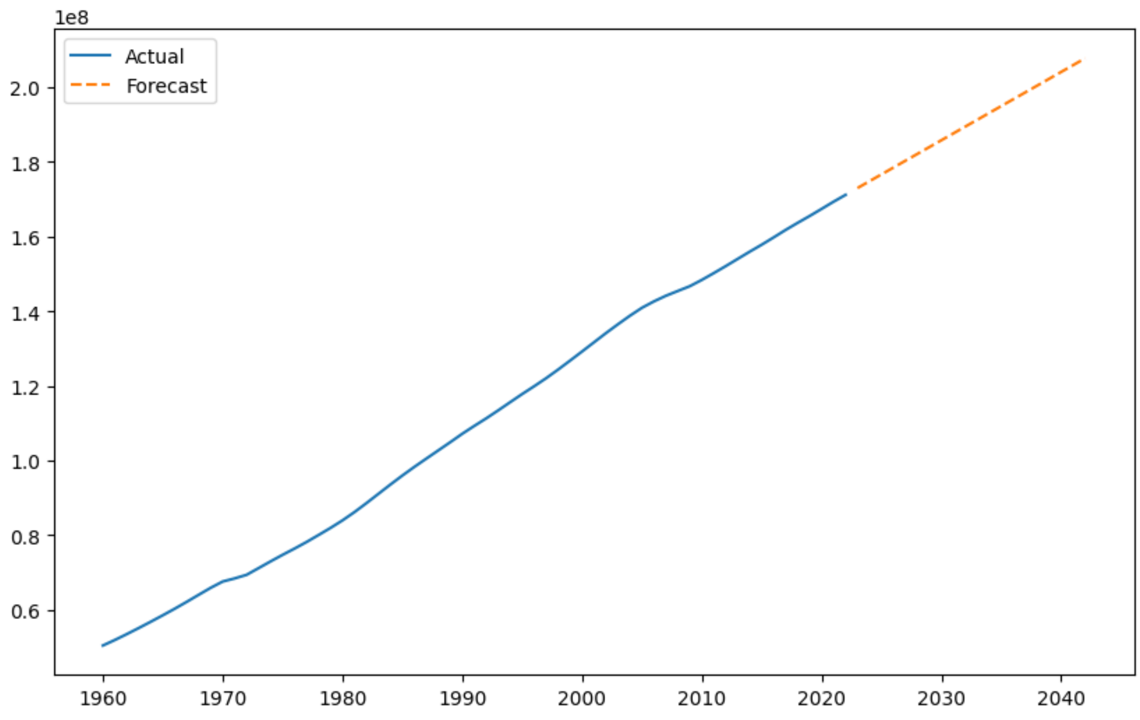

Figure 4.2.5: state space model forecast

As depicted in the above figure, the temporal consistency of the model is brilliant, it shows generalization on a very fundamental level, the State Space Model showcased a MAE of 227,678.52 and an RMSE of 248,678.73. These metrics underscore its competence in capturing temporal patterns in population data. Its sensitivity to seasonality and trends makes it a suitable choice for applications where understanding long-term population trends is essential.

Holt-Winters: Consistency and Robustness

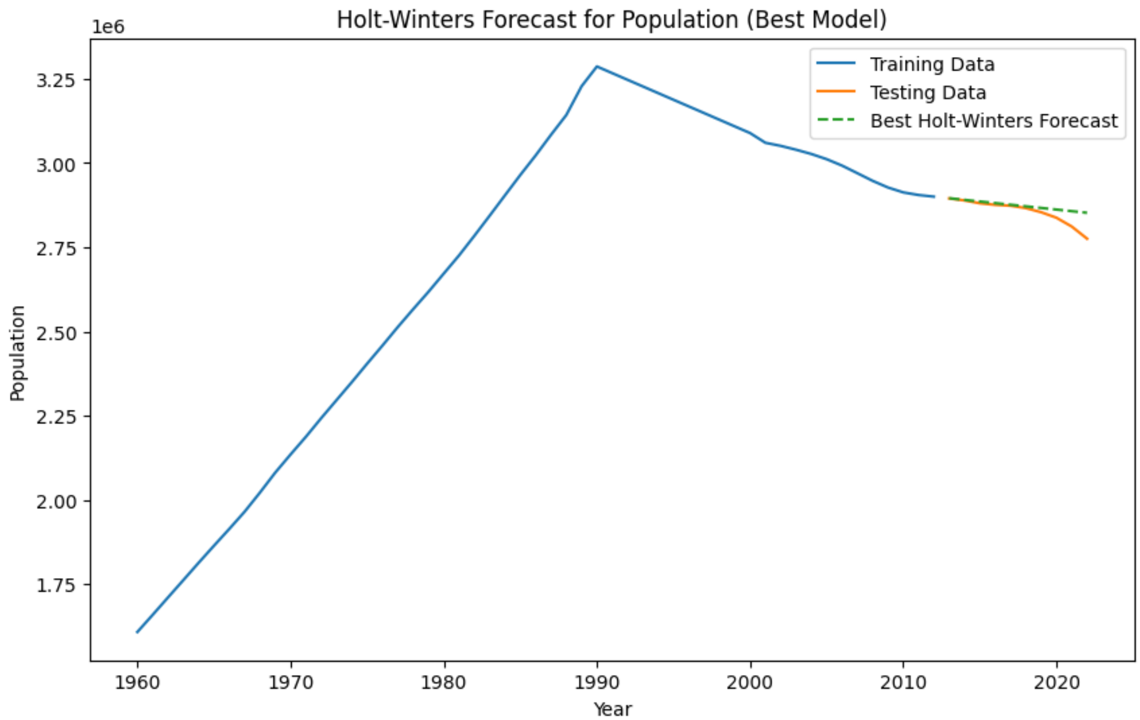

Figure 4.2.6: holt winters forecast. Showing adaption to challenging scenarios

Holt-Winters, As depicted in Figure 4, the Holt-Winters model demonstrates a robust performance in population prediction on a country dataset. With a Mean Absolute Error (MAE) of 18,016.03 and a Root Mean Square Error (RMSE) of 29,704.29, the model exhibits minor deviations from the actual values. However, these are offset by its reliability and consistency in forecasting, making it a dependable choice for scenarios requiring steady population predictions. The plot in Figure 4 effectively illustrates the model's performance, underscoring its potential for reliable population forecasting.

SARIMA: Navigating Complexity

In order to maintain an uninterrupted and unbiased opinion on the whole process, we tried to work with the US data for this model primarily.

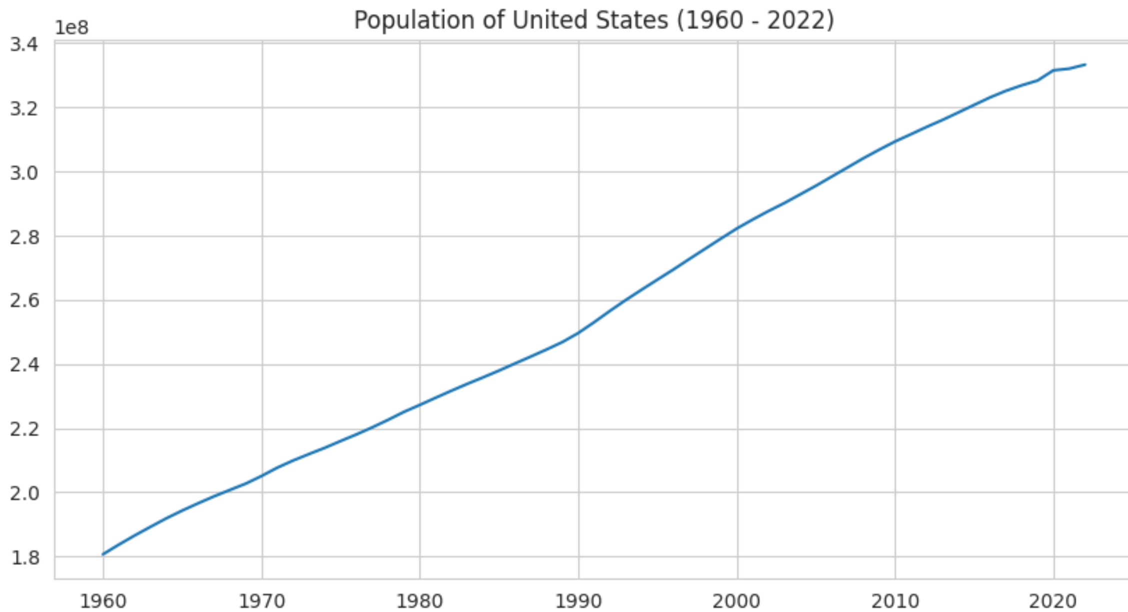

Figure 4.2.7: Raw population data for the USA

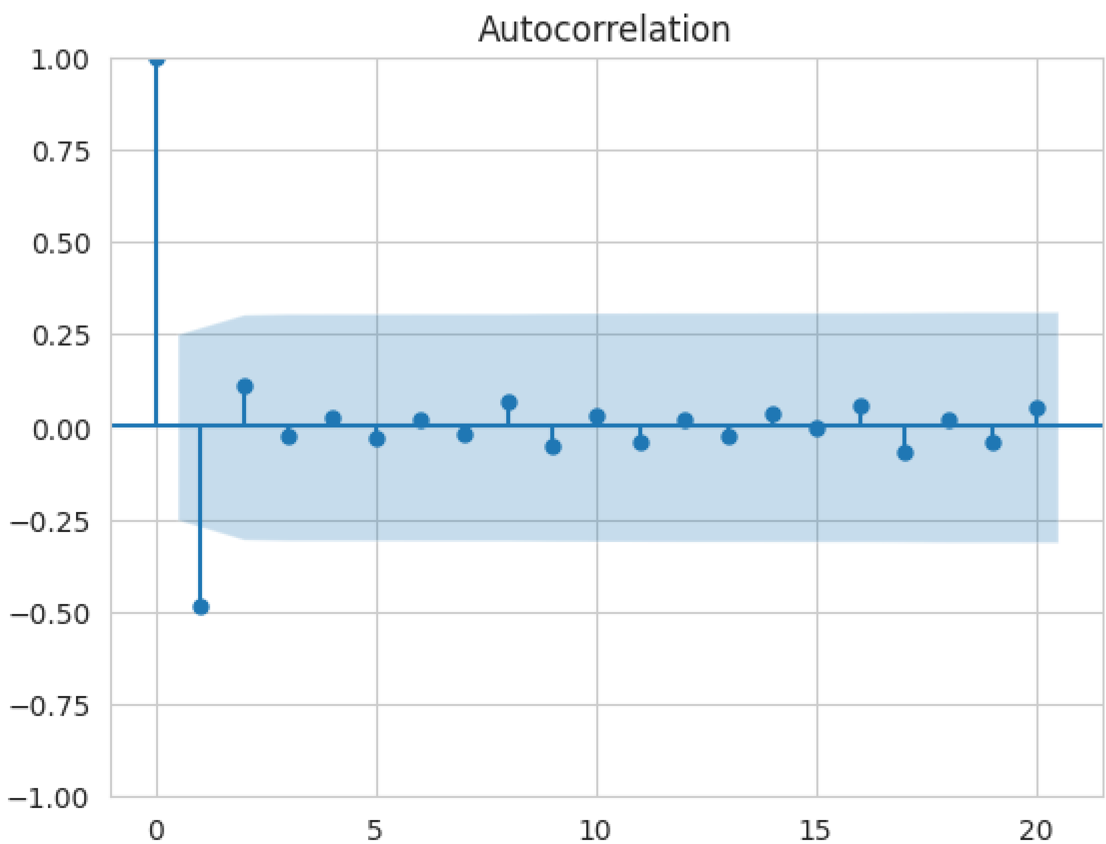

Figure 4.2.8: autocorrelation plot for SARIMA model

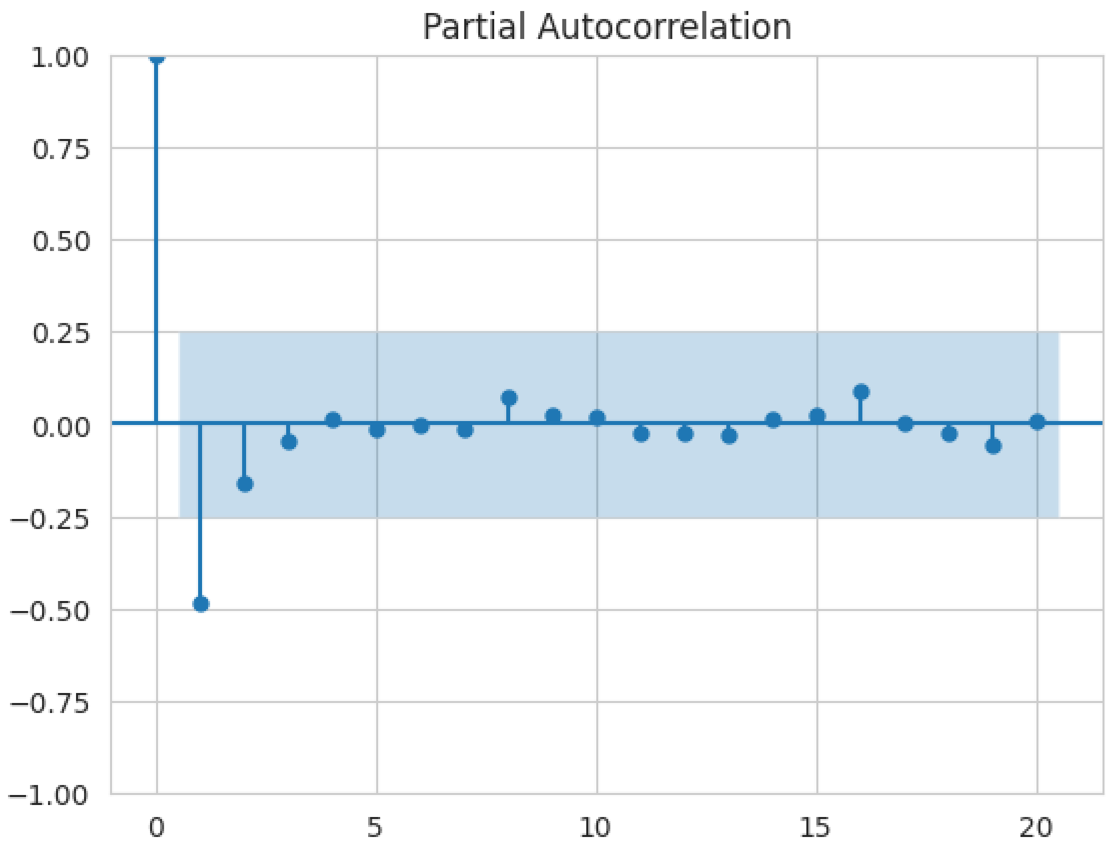

Figure 4.2.9: partial autocorrelation plot for SARIMA.

The population data of the United States exhibits a pattern of decline, suggesting that the current value is influenced by its recent past values. There are notable correlations at lag 1 and 2 in the partial autocorrelation plot, indicating direct influences from these time periods. After taking the second difference, the series becomes stationary according to the results of the Augmented Dickey-Fuller (ADF) test. Using the second difference of the population data could potentially enhance the accuracy of the SARIMA model for weather forecasting, based on these insights. The insights are further refined by the partial autocorrelation plot, which highlights the direct influences at different lags. This is crucial for selecting the parameters for the SARIMA model.

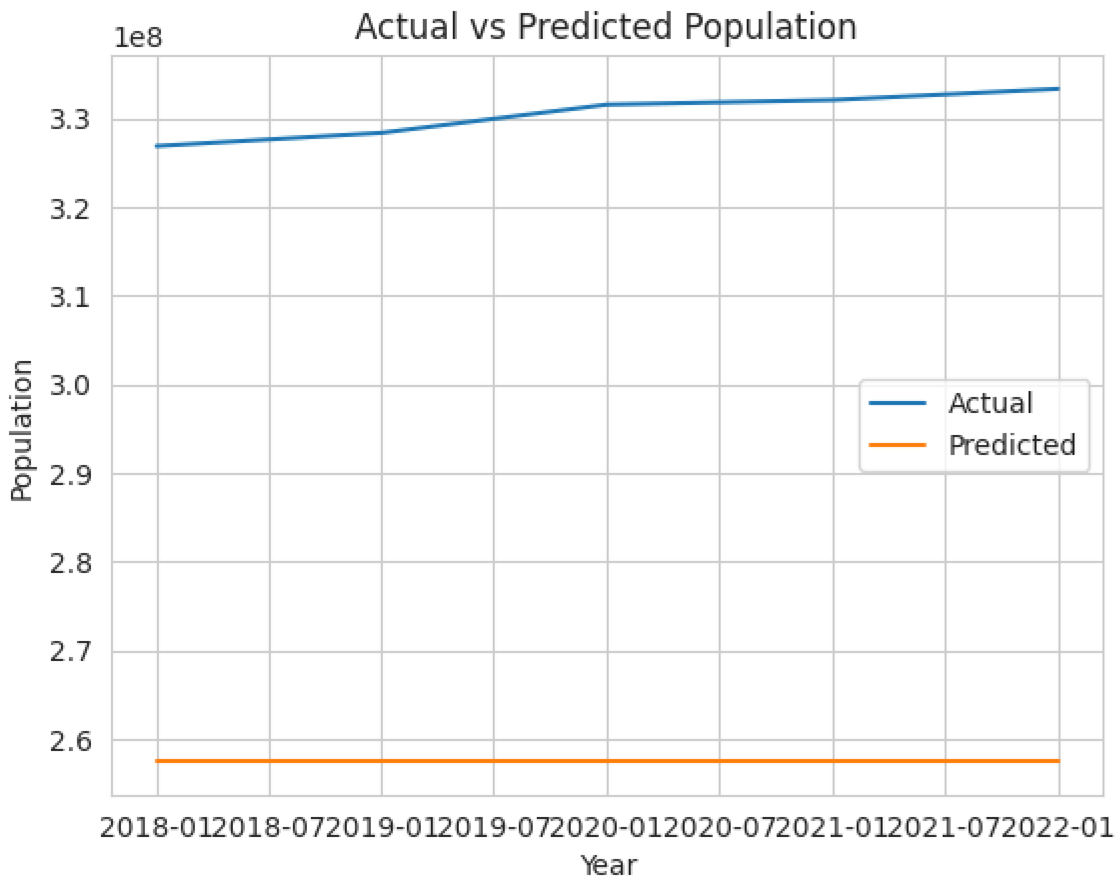

Figure 4.2.10: actual vs predicted value close up view for SARIMA.

The population forecasting is effectively handled by the SARIMA model, which achieves an RMSE of around 79,193,532.98. The plot provided shows how the model's predictions consistently overestimate population growth when compared to the actual data. In spite of its somewhat high RMSE, the model's capacity to deal with complex temporal relationships in the data provides valuable insights into subtle population trends. The plot demonstrates the capability of SARIMA in population forecasting research, highlighting the challenges it may pose for other models.

4.3 Discussion

Policy and Planning: Providing policymakers and urban planners with accurate population forecasts allows for data-driven insights, leading to efficient resource allocation, infrastructure planning, and informed decision-making.

Population forecasts can be used by businesses and industries to strategically expand, create jobs, and customize products/services for specific demographic segments, thus contributing to economic development. Designing environmentally sustainable cities and regions is aided by precise population predictions, which help minimize ecological footprints. Proactive planning in healthcare and education is facilitated by population forecasts, which help ensure access to quality services. During crises, having reliable population data is crucial for emergency management. It greatly improves preparedness and response, ultimately saving lives. Population forecasts are essential for developing long-term visions that aim to achieve sustainable growth and enhance quality of life. Our research findings highlight the promising capabilities of machine learning models in accurately predicting and providing valuable insights into population forecasting. There are different strengths in each model that can be customized to meet specific forecasting requirements. These findings can be used by policymakers, urban planners, businesses, and researchers to inform their decision-making and contribute to a more prosperous and sustainable future. The models demonstrate precision, robustness, and temporal sensitivity, which create opportunities for additional research and application. Our work in the field of population forecasting continues to evolve, providing valuable tools and insights to inform decisions that shape communities and societies.

CHAPTER 5

IMPACT ON SOCIETY ENVIRONMENT AND SUSTAINABILITY

5.1 Impact on Society

Population forecasting has a profound impact on society, influencing economic, social, and geopolitical aspects across the globe. The accuracy of such forecasts plays a crucial role in shaping policy and planning decisions. For instance, historical projections by the United Nations, dating back to 1968, have shown reasonable accuracy for global population estimates, albeit with larger errors in earlier decades. These projections, while generally close to real figures in the short-to-medium term, tend to diverge over longer timescales due to the complexities of predicting changes in fertility rates and life expectancy. Errors in these projections can significantly affect regions differently, often due to initial inaccuracies in population data, especially in countries with poor census records. For example, in Europe, projections often overestimated the number of future children while underestimating the elderly population, due to incorrect assumptions about birth rates and life expectancy improvements. Such inaccuracies have implications for planning in areas such as healthcare, education, and social services, highlighting the importance of accurate and reliable population forecasting in societal planning and development.

5.2 Impact on Environment

Population growth and its forecasting have direct implications for environmental sustainability. The scale and pace of population growth can exacerbate environmental pressures, such as increased demand for natural resources, higher levels of pollution, and greater strain on ecosystems. Accurate population forecasts are vital for environmental planning and management, helping to anticipate and mitigate these impacts. They are essential for developing sustainable practices in areas such as urban development, resource allocation, and environmental conservation.

5.3 Ethical Aspects

Ethical considerations in population forecasting include issues related to privacy, data accuracy, and the potential misuse of demographic data. Ensuring the confidentiality and ethical use of personal data collected for forecasting purposes is paramount. Additionally, the ethical implications of using demographic data for policy-making, which may disproportionately affect certain groups within society, must be carefully considered. This involves balancing the need for accurate data for effective planning against the rights and interests of individuals and communities.

5.4 Sustainability Plan

A sustainability plan in the context of population forecasting involves developing strategies that ensure the long-term viability and effectiveness of forecasting efforts. This includes continuous improvement in data collection methods, incorporation of emerging technologies, and adaptation to changing demographic trends. It also involves collaboration with various stakeholders, including governments, international organizations, and local communities, to ensure that forecasting informs sustainable development policies and practices. The plan should emphasize the responsible use of resources, promote environmentally friendly practices, and consider the long-term impact of population changes on both the environment and society.

CHAPTER 6

SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH

6.1 Summary of the Study

Our research aims to utilize machine learning to predict population trends. We have thoroughly evaluated various models such as Facebook Prophet, LSTM, State Space Model, Holt-Winters, and SARIMA. This evaluation has provided valuable insights into their capabilities and the implications they have for accurate population predictions.

6.2 Conclusions

The spectrum of achievements we have uncovered reveals how each model brings its own unique strengths to the field of population forecasting. Facebook Prophet's exceptional performance, boasting an impressive Mean Absolute Percentage Error (MAPE) of 0.48%, highlights the potential for accurate and dependable population predictions. Our forecasting toolkit is enriched by the collective capabilities of LSTM in capturing population trends, the State Space Model's temporal sensitivity, Holt-Winters' consistency, and SARIMA's capacity to navigate complex temporal dependencies. The foundation of well-informed urban planning lies in accurate population forecasts. Our models can assist city officials and urban planners in predicting population growth, which in turn helps guide the development of sustainable infrastructure, transportation networks, and public services. As a result, it promotes the creation of urban environments that are more livable, efficient, and vibrant. Healthcare providers and policymakers can use population forecasts to predict healthcare demands, allocate resources efficiently, and guarantee access to medical services. Through our research, individuals are equipped to make informed choices that improve public health and the delivery of healthcare. Utilize our population forecasts to strategically target areas with growing populations, benefiting businesses and industries poised for expansion. By aligning their efforts with current demographic trends, they have the potential to boost economic growth and generate employment prospects.

6.3 Implication for Further Study

As we wrap up this research, we recognize that the future holds potential and fascination. The discoveries we made open up possibilities for future research to improve and broaden the use of machine learning in predicting population trends. Possible areas to explore could be:
By combining the strengths of multiple models, more accurate and robust forecasts can be achieved with hybrid models. Exploring the fusion of machine learning and traditional statistical approaches is worth considering.

Refinement: By continuously adjusting model parameters and refining data preprocessing techniques, the forecasting capabilities can be improved, leading to a potential reduction in minor deviations. By incorporating scenario analysis, population forecasts can become more adaptive and resilient, as they account for unforeseen events or policy changes.

Enhancing Collaboration: Working together with specialists from various fields, including demographers, epidemiologists, and environmental scientists, can offer fresh perspectives and enhance the applicability of forecasts.

Enhancing Data: Having access to more detailed and diverse data sources, such as social and environmental factors, can greatly improve the accuracy and comprehensiveness of population forecasts.

We are dedicated to pushing the boundaries of population forecasting through the use of machine learning. Our findings have far-reaching implications for urban planning, healthcare, and economic development. They provide decision-makers with a valuable toolbox for making data-driven decisions. Looking forward, our top priority is to enhance accuracy and adaptability. We are committed to ensuring that the models and insights we present here continue to make a positive impact on our communities and societies.

REFERENCES

[1] Şahinarslan, Fatih Veli, Ahmet Tezcan Tekin, and Ferhan Çebi. "Application of machine learning algorithms for population forecasting." *International Journal of Data Science* 6, no. 4 (2021): 257-270.

[2] "Population Prediction Using Machine learning", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.6, Issue 6, page no.189-192, June 2019, Available:http://www.jetir.org/papers/JETIRDC06033.pdf

[3] Moieni, Rezza, Peter Mousaferiadis, and Leyla Roohi. "A Study on Diversity Prediction with Machine Learning and Small Data." *Open Journal of Social Sciences* 11, no. 2 (2023): 18-31.

[4] Grossman, Irina, Tom Wilson, and Jeromey Temple. "Forecasting small area populations with long short-term memory networks." *Socio-Economic Planning Sciences* (2023): 101658.

[5] N. Ashioba and N. N. Daniel, "Population Forecasting System Using Machine Learning Algorithm," International Journal of Computer Trends and Technology, vol. 68, no. 12, pp. 40-43, Dec. 2020. [Online]. Available: https://doi.org/10.14445/22312803/IJCTT-V68I12P109.

[6] Bandyopadhyay, Goutami, and Surajit Chattopadhyay. "An Artificial Neural Net approach to forecast the population of India." arXiv preprint nlin/0607058 (2006).

[7] Şahinarslan, Fatih Veli, Ahmet Tezcan Tekin, and Ferhan Çebi. "Application of machine learning algorithms for population forecasting." *International Journal of Data Science* 6, no. 4 (2021): 257-270.

[8] Wilson, Tom, Irina Grossman, Monica Alexander, Phil Rees, and Jeromey Temple. "Methods for small area population forecasts: State-of-the-art and research needs." *Population Research and Policy Review* 41, no. 3 (2022): 865-898.

[9] Baker, Jack, David Swanson, and Jeff Tayman. "Boosted Regression Trees for Small-Area Population Forecasting." *Population Research and Policy Review* 42, no. 4 (2023): 51.

[10] Wilson, Tom, Irina Grossman, and Jeromey Temple. "Evaluation of the best M4 competition methods for small area population forecasting." *International Journal of Forecasting* (2021).

[11] Chen, Yimin, Xia Li, Kangning Huang, Ming Luo, and Minyi Gao. "High-resolution gridded population projections for China under the shared socioeconomic pathways." *Earth's Future* 8, no. 6 (2020): e2020EF001491.

[12] Wu, Chuhan, Fangzhao Wu, Junxin Liu, Shaojian He, Yongfeng Huang, and Xing Xie. "Neural demographic prediction using search query." In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, pp. 654-662. 2019.

[13] Swanson, David A., Jeff Tayman, and T. M. Bryan. "MAPE-R: a rescaled measure of accuracy for cross-sectional subnational population forecasts." *Journal of Population Research* 28 (2011): 225-243.

[14] Wang, Chian-Yue, and Shin-Jye Lee. "Regional population forecast and analysis based on machine learning strategy." *Entropy* 23, no. 6 (2021): 656.

[15] Li, Dong, Yanyan Yu, and Bo Wang. "Urban population prediction based on multi-objective lioness optimization algorithm and system dynamics model." *Scientific Reports* 13, no. 1 (2023): 11836.

[16] Büchler, Simon, Dongxiao Niu, and Anne Kinsella Thompson. "Predicting urban growth with machine learning." *MIT Center for Real Estate Research Paper* 21/06 (2021).

[17] Şahinarslan, Fatih Veli, Ahmet Tezcan Tekin, and Ferhan Çebi. "Machine learning algorithms to forecast population: Turkey example." *İstanbul Technical University, Internatıonal Engıneerıng And Technology Management* (2019).

[18] Juan, Li, and Wei-jie Liu. "Population forecasting in China based on the Grey-Markov model." In *2011 International Conference on Information Management, Innovation Management and Industrial Engineering*, vol. 3, pp. 133-136. IEEE, 2011.

[19] Bettiza, Martaleli. "Forecasting Total Population Using Chen, Cheng, and Markov Chain Fuzzy Time Series Models." In *2020 12th International Conference on Information Technology and Electrical Engineering (ICITEE)*, pp. 135-140. IEEE, 2020.

[20] Ilyas, Muhammad, Shaheen Abbas, Wajid Ali, and S. Masood Raza. "Application of Stochastic Regression Models: ARIMA (p, d, q)-HW Algorithm Approach for Human Population Forecasting." *ratio* 1: 5.

[21] Hussain, Jassim N. "A comparative study to choose the appropriate growth model to forecast COVID-19 cases in Iraq." In *Journal of Physics: Conference Series*, vol. 2322, no. 1, p. 012026. IOP Publishing, 2022.

[22] Papastefanopoulos, Vasilis, Pantelis Linardatos, and Sotiris Kotsiantis. "COVID-19: a comparison of time series methods to forecast percentage of active cases per population." *Applied sciences* 10, no. 11 (2020): 3880.

[23] Sangrody, Hossein, Ning Zhou, Salih Tutun, Benyamin Khorramdel, Mahdi Motalleb, and Morteza Sarailoo. "Long term forecasting using machine learning methods." In *2018 IEEE Power and Energy Conference at Illinois (PECI)*, pp. 1-5. IEEE, 2018.

[24] Ronald Doni, A., T. Sasi Praba, and S. Murugan. "Weather and population based forecasting of novel COVID-19 using deep learning approaches." *International Journal of System Assurance Engineering and Management* 13, no. Suppl 1 (2022): 100-110.

[25] Zhang, Yunong, Wan Li, Binbin Qiu, Hongzhou Tan, and Ziyi Luo. "UK population forecast using twice-pruning Chebyshev-polynomial WASD neuronet." In *2016 Chinese Control and Decision Conference (CCDC)*, pp. 3029-3034. IEEE, 2016.

[26] Wan, Yi, Lei Xiao, and Chengwen Wu. "An optimum intelligent algorithm and its application in population statistic and forecast." In *2009 WRI Global Congress on Intelligent Systems*, vol. 4, pp. 40-44. IEEE, 2009.

[27] Ying, Jiang Mei, Xiong Li Ran, Hu Zhi Ding, and Niu Le De. "An Analysis of the Change in Population Size of Chinese Wa Nationality and Predictive Based on Census Data and Spectrum Population Prediction Software." In *2014 Sixth International Conference on Measuring Technology and Mechatronics Automation*, pp. 75-77. IEEE, 2014.

[28] Chen, Han-Jun, and Dong-Wei Huang. "Chaos Analysis and Modeling for Predicting Long-term Population." In *2009 Second International Conference on Information and Computing Science*, vol. 3, pp. 357-359. IEEE, 2009.

[29] Abbasov, A. M., and M. H. Mamedova. "Application of fuzzy time series to population forecasting." *Vienna University of Technology* 12 (2003): 545-552.

[30] Hu, Ping, and Shun-rong Yan. "The Comparative Study of Wuhan Population Prediction Models." In *2010 International Conference on Challenges in Environmental Science and Computer Engineering*, vol. 1, pp. 22-25. IEEE, 2010.

[31] Zhang, Yunong, Jianxi Liu, Xiaogang Yan, Binbin Qiu, and Tianjian Qiao. "WASD neuronet prediction for China's population." In *2015 IEEE International Conference on Information and Automation*, pp. 797-802. IEEE, 2015.

[32] Xiaobo, Yang, Chen Bing, Chen Chuxiang, and Wang Yuncheng. "Dynamic model of population prediction based on Aging Chains." In *2016 Chinese Control and Decision Conference (CCDC)*, pp. 216-219. IEEE, 2016.

[33] Zhang, Yunong, Penghao He, Jianfeng Wen, Liangyu He, and Hongzhou Tan. "China population prediction with the aid of fast, optimal and accurate WASD neuronet compared with BP neuronet." In *2016 35th Chinese Control Conference (CCC)*, pp. 3673-3678. IEEE, 2016.

[34] Li, Bin, Tianfei Wang, and Liping Jia. "Application of improved logarithm logistic models in population prediction." In *2012 Eighth International Conference on Computational Intelligence and Security*, pp. 99-102. IEEE, 2012.

[35] Zhang, Yunong, Jinjin Wang, Qingkai Zeng, Heng Qiu, and Hongzhou Tan. "Near future prediction of European population through Chebyshev-activation WASD neuronet." In *2015 Sixth International Conference on Intelligent Control and Information Processing (ICICIP)*, pp. 134-139. IEEE, 2015.

[36] Zhu, Mingzhang, Changchun Pan, Genke Yang, and Haichun Liu. "Prediction of population distribution on campus based on historical location data." In *2016 Chinese Control and Decision Conference (CCDC)*, pp. 2849-2854. IEEE, 2016.

APPENDIX

The main novelty of this research lies in its unique approach. Global population data and it's forecasting rarely utilize time series analysis. Time series models are typically employed within specific local demographics to address specific tasks. These models take into account other relevant features and have shown promising results on a global scale.

Plagiarism Report

world population detection using times series analysis and machine learning.

ORIGINALITY REPORT

22%	18%	11%	14%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	5%
2	www.mdpi.com Internet Source	1%
3	www.researchgate.net Internet Source	1%