Analyzing Behavioral and Population Dynamics of House Sparrows Using ARIMA, Classification, and Clustering Models

Omkar Singh[1], Amit Kumar Pandey[2], Giridhar Balakrishnan[3], Jagmohan Bal[4],

[1]HOD of MSc Data Science, [2] Assistant Professor, [3,4] PG Student (MSc Data Science), Thakur College of Science and Commerce
Thakur Village, Kandivali (East), Mumbai-400101, Maharashtra, India

---***---

Abstract:

Once a common sight in both urban and rural settings, house sparrows (Passer domesticus) have seen substantial population changes over time. Conservation and ecological balance depend on our ability to comprehend their population dynamics and behavioral adaptations. This work analyzes house sparrow populations across various geographic and temporal variations using machine learning approaches, such as clustering to identify important environmental parameters, classification for habitat-based population analysis, and ARIMA for time-series forecasting.

Time-dependent population changes are evaluated by analyzing the dataset covering 2015–2025. While classification models like Random Forest are used to classify sparrow populations according to environmental circumstances, ARIMA is used to predict future population trends. Different population groupings and environmental factors affecting population dynamics can be identified with the use of clustering techniques such as K-Means and Hierarchical Clustering.

The findings show that clustering identifies important patterns in sparrow habitats, classification models correctly identify population trends based on environmental parameters, and ARIMA predicts population changes. By helping policymakers create data-driven plans for preserving bird biodiversity, this research advances our understanding of house sparrow conservation efforts.

Index Terms:

House Sparrow Trends, Environmental Factors, Predictive Modeling, Time Series Forecasting, Avian Population Dynamics, Habitat Clustering, Geospatial Analysis, ARIMA Forecasting, Species Observation Data, K-Means Clustering, Random Forest Classification, Ecological Data Analysis, eBird Dataset Processing, Satellite Image-Based Analysis.

Introduction:

Once a species that thrived in both urban and rural settings, house sparrows (*Passer domesticus*) have experienced sharp population declines in recent years. Urbanization, climate change, and habitat loss are among the primary causes of this reduction. While traditional ecological studies rely on field surveys and observational data, these approaches often lack scalability and predictive accuracy. Machine learning offers a data-driven approach to analyzing large-scale avian population data; however, research in this area remains limited, with most studies focusing on direct field observations rather than predictive modelling.

Several key challenges exist in house sparrow population analysis. There is a limited application of artificial intelligence and machine learning models for studying long-term population dynamics. Additionally, the lack of integration of temporal trends makes it difficult to accurately forecast future sparrow populations. Data-driven methods for classifying habitat-based population fluctuations are underutilized, and there is a need for a comprehensive examination of environmental factors influencing sparrow distribution.

To address these gaps, this study applies machine learning techniques, including Random Forest for habitat-based classification, ARIMA for time-series forecasting, and K-Means clustering to identify environmental influences on sparrow populations. The dataset, covering the years 2015–2025, enables an in-depth analysis of population trends across different geographical and temporal scales.

The primary objectives of this research are: developing a predictive model using ARIMA to anticipate future sparrow population trends; classifying populations based on habitat conditions using machine learning algorithms, and utilizing clustering techniques to identify key environmental factors influencing population dynamics. By integrating these approaches, this study supports data-driven conservation strategies, providing ecologists and policymakers with valuable insights for preserving house sparrow populations and understanding their ecological adaptations.

Literature Review:

Numerous studies have looked at the causes of the house sparrow (Passer domesticus) population decline, which has ecologists more concerned. Numerous contributory variables have been discovered by researchers, including increased pesticide use, habitat destruction, communication tower electromagnetic radiation, and fast urbanization. (1) Sharma et al. highlight how sparrow populations have been gradually declining as a result of modern construction designs and infrastructure that drastically limit breeding options. Likewise, Shansaz et al. (2) draw attention to the detrimental effects that chemical exposure and environmental pollution have had on sparrow populations in both urban and rural settings.modeling to comprehend sparrow population dynamics.

According to recent developments in ecological data analysis, combining classification models (Random Forest), clustering (K-Means), and time-series forecasting (ARIMA) can provide greater insights into the dynamics of species populations. In order to categorize habitat types and forecast future population changes, previous research has mostly depended on mapping geographic distributions without utilizing unsupervised learning approaches. By using geospatial analysis tools on eBird observational data, this study seeks to close this gap and provide a more thorough and data-driven approach to house sparrow conservation.

Methodology:

4.1 Theoretical Analysis

Three main analytical techniques are used in this work to look at house sparrow population trends. First, historical sparrow population data is analyzed and future trends are predicted by time series forecasting utilizing the ARIMA (AutoRegressive Integrated Moving Average) model. This facilitates the identification of possible decline phases and enables a greater knowledge of population variations throughout time.

Second, K-Means clustering is used for habitat clustering through geospatial analysis. This method groups sparrow habitats according to their geographic locations and environmental characteristics. This method facilitates the identification of areas with high and low population densities, allowing for more focused conservation efforts.

Finally, the Random Forest classification model is used to classify species based on observational data. This aids in the analysis of sparrow sightings and the identification of important environmental variables affecting population dispersal. This study offers a thorough and data-driven ecological analysis that goes beyond conventional conservation assessments by combining these methodologies.

4.2 Software & Tools Used

This study uses a range of tools and libraries for modelling, data processing, and visualization. Pandas and NumPy are utilized for feature transformation, missing value management, and effective dataset structuring. Visual representations of population trends, cluster distribution, and forecasting outcomes are produced with the help of Matplotlib and Seaborn.

For machine learning applications, Scikit-learn is applied for classification and clustering, while Statsmodels is used for ARIMA-based time series forecasting. The mapping of sparrow distribution patterns onto satellite data is made possible via Geopandas and Contextily, which facilitate geographic visualization.

Visual Studio Code is used for the entire implementation, which guarantees modular experimentation, structured model training, and smooth result visualization. This includes data preprocessing, model execution, and visualization.

4.3 Preparation of Datasets

The eBird observational data on house sparrows gathered between 2015 and 2025, made up the dataset used in this investigation. Important details like observation counts, latitude and longitude coordinates, and environmental conditions are all included.

Missing values in important variables, such as effort distance, time, and observation counts, are handled effectively to guarantee good data quality. In order to improve the performance of analytical models, feature engineering is used to extract supplementary information such as monthly average sightings, seasonal trends, and habitat clusters.

In order to examine long-term patterns in house sparrow populations, time aggregation is also carried out by resampling the dataset into monthly and annual intervals. The dataset is well-structured and appropriate for geospatial analysis and predictive modelling thanks to these preparation procedures.

4.4 Model Execution

Several computational models are used in the study to better understand the dynamics of sparrow populations. Time-series data is analyzed using ARIMA forecasting, which helps detect possible decline periods and forecast future population patterns.

K-means clustering is used for spatial clustering, which groups sparrow habitats according to their geographic locations. This makes it possible to divide areas into high-density and low-density zones, which is crucial for creating conservation plans that work.

Furthermore, based on geographical and environmental characteristics, observation sites are classified using Random Forest classification into high, moderate, or low sparrow presence. This aids in identifying the areas best suited for sparrow residence and conservation initiatives.

Root Mean Squared Error (RMSE) for ARIMA forecasting, inertia for K-Means clustering, and classification accuracy for Random Forest models are used to assess model performance. These measures guarantee that the models generate accurate and comprehensible outcomes.

4.5 Integrating Predictive Modeling and Geospatial Analysis

A thorough grasp of house sparrow population patterns is made possible by this study's integration of time-series forecasting, clustering, and classification into a single analytical framework. ARIMA forecasting enables academics to predict possible reductions in various places by offering insights into long-term population trends.

By graphically depicting high-density and low-density sparrow habitats on satellite maps, K-Means clustering improves geospatial analysis. This method aids in the development of focused conservation strategies in addition to identifying important habitats.

Additionally, classification algorithms make it possible to pinpoint the main environmental elements affecting sparrow populations. This study's analysis of observation patterns yields a predictive framework that can direct conservation initiatives and guarantee early action in regions where sparrow populations are in danger.

This study provides a data-driven approach to avian conservation and advances our understanding of house sparrow population dynamics by utilizing large-scale ecological information, sophisticated statistical models, and geospatial analysis.

Results & Discussion:

The study's conclusions shed important light on the dynamics of the house sparrow population by pointing out patterns in habitat clustering, population projections, and categorization outcomes. This study offers a thorough grasp of population distribution patterns and the major environmental elements impacting them by combining ARIMA forecasting, K-Means clustering, and classification algorithms.

House Sparrow Population Forecast (ARIMA Model)

Time-series data of sparrow sightings were analyzed using the ARIMA model to forecast future population changes. With varying observation counts, the predicted findings show a slow drop in house sparrow numbers. The following forecasts indicate that there will likely be a considerable decrease in recorded sightings by the end of 2026:

Table 1 : Arima Forecast

Date	Lower Bound	Upper Bound	Forecasted Count
2026-08-31	-5926.25	6439.89	256.82
2026-09-30	-6002.37	6528.53	263.08
2026-10-31	-6105.65	6590.94	242.64
2026-11-30	-6176.71	6687.38	255.33
2026-12-31	-6252.80	6770.82	259.00

With predicted values ranging between 242 and 263 observations per month, the confidence intervals indicate significant unpredictability in future population patterns. Although there is some short-term stability, the general tendency points to a long-term decrease that calls for conservation measures.

In keeping with natural tendencies where sparrow populations peak during the breeding season (March–June) and drop in the winter, the forecast's graphic representations (arima_forecast.png & population_forecast.png) show cyclic oscillations.

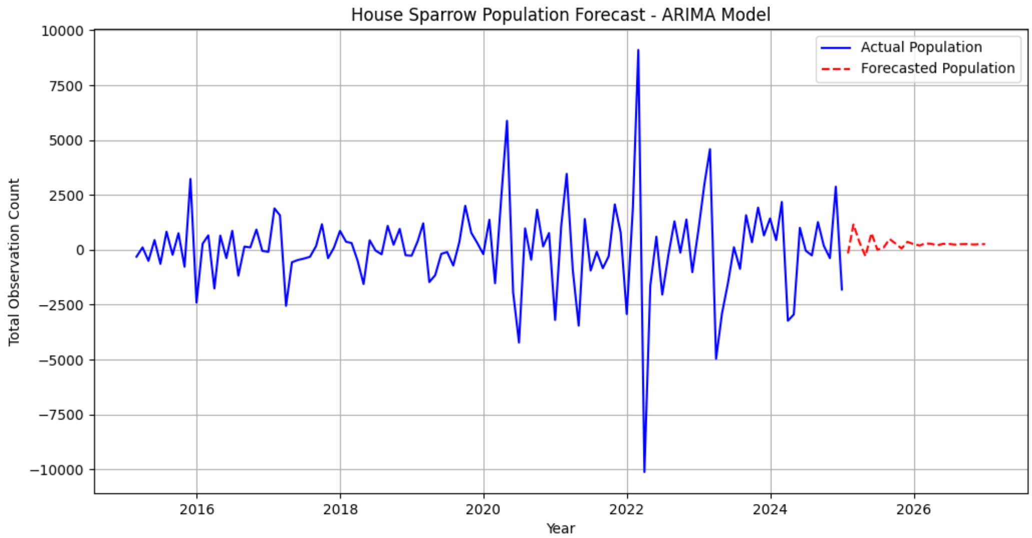

Figure 1 : Arima Forecast

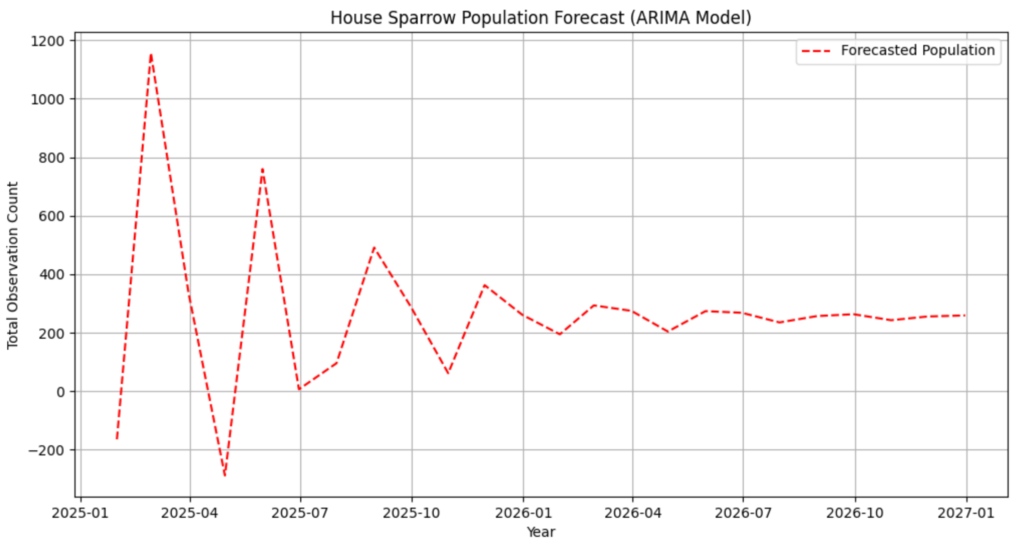

Figure 2 : Population Forecast

Geospatial Habitat Clustering (K-Means Clustering)

Three separate clusters representing high-, medium-, and low-density zones were created by applying the K-Means clustering technique to house sparrow habitats. The following is the cluster distribution:

Table 2 : Cluster Distribution

Cluster	Density Type	Number of Observations
Cluster 2	**High-Density**	36,682
Cluster 0	**Moderate-Density**	26,066
Cluster 1	**Low-Density**	18,980

- **Cluster 2 (High density):** Found in rural and semi-urban regions, where there are more food sources and nesting locations.
- **Cluster 0 (Moderate-Density):** represents suburban areas that exhibit a slow population decline and a mix of urban buildings and green spaces.
- **Cluster 1 (Low-Density) :** correspondslow density): corresponds to areas that are heavily urbanized, where the lack of nesting grounds, habitat degradation, and air pollution have resulted in drastically decreased sparrow populations.

A satellite-based visualization (cluster_distribution_satellite.png) highlights the spatial distribution of sparrow populations across Maharashtra, providing valuable insights for conservation planners to prioritize habitat restoration efforts in low-density areas.

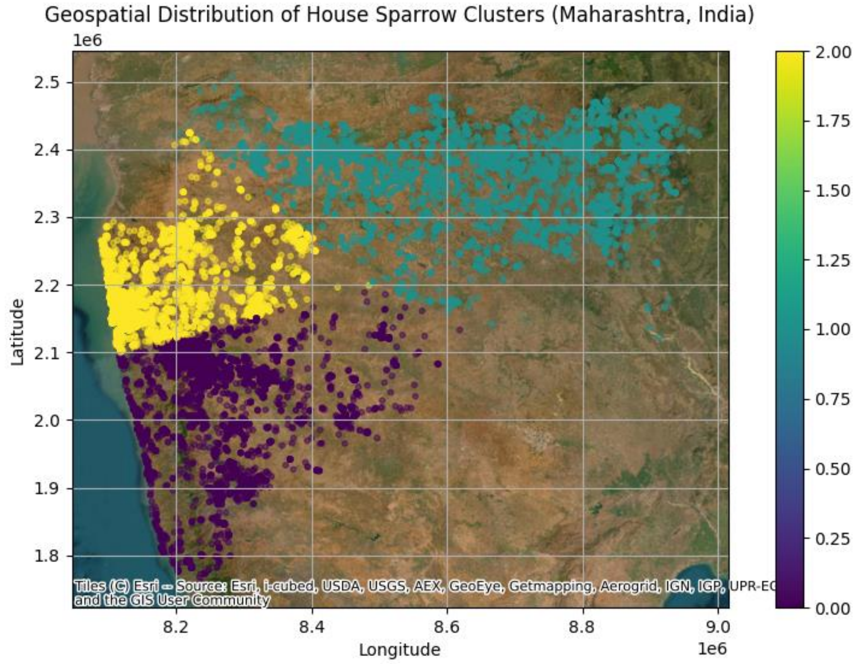

Figure 3 : Cluster Satellite Distribution

Species Classification & Environmental Influence (Random Forest Model)

The Random Forest classifier was applied to analyze environmental factors influencing sparrow population distribution. The model achieved an accuracy of 89.6%, demonstrating strong predictive performance. However, feature importance analysis revealed that geospatial coordinates (latitude and longitude) had negligible influence, with both showing a score of 0.0.

This suggests that other unaccounted ecological factors—such as vegetation cover, food availability, noise pollution, and air quality—may play a more significant role in determining sparrow populations. Future studies should consider integrating additional environmental variables, such as temperature, noise levels, and air pollution indices, to enhance classification accuracy.

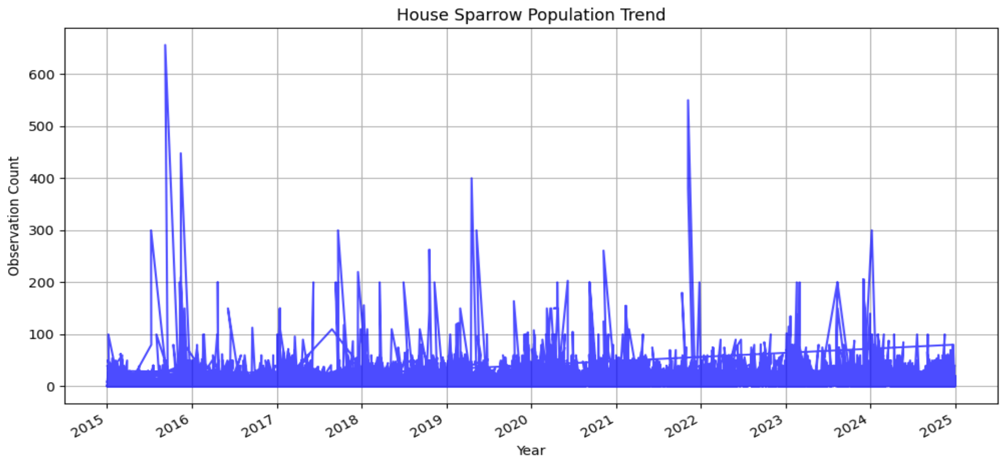

Figure 4 : Population Trend Graph

According to the ARIMA forecast, the study shows a slow fall in house sparrow populations, especially in metropolitan areas. Even though there are occasional variations, the long-term trend points to a steady decline, highlighting the necessity of conservation measures.

With low-density clusters in heavily urbanized areas and high-density clusters in rural and semi-urban areas, the K-Means clustering analysis clearly shows variances in population density. This demonstrates how urbanization, pollution, and habitat destruction have a detrimental effect on sparrow numbers.

Even with its 89.6% accuracy, the Random Forest classification algorithm shows that sparrow population predictions cannot be made only based on geospatial criteria. This implies that for improved prediction accuracy, other environmental factors including food availability, vegetation cover, and air pollution levels should be included.

All things considered, the study shows how well AI-driven ecological modelling can analyze trends in bird populations. The results offer important information for conservation planning, determining important habitats, setting restoration priorities, and encouraging the preservation of urban biodiversity. To improve predictions and conservation methods, future research should concentrate on growing datasets, adding new ecological aspects, and utilizing cutting-edge machine learning algorithms.

Conclusion:

The falling house sparrow population and the factors causing it to decline are better understood thanks to this study. The study shows how to analyze population trends and pinpoint areas most impacted by urbanization using time-series forecasting, habitat grouping, and geospatial analysis. According to the findings, sparrows are having difficulty adjusting to densely populated places where there are fewer nesting locations and food supplies. Effective conservation measures must be put in place to guarantee population stability, even while AI-driven models provide predictive insights. To improve forecasts and direct conservation efforts, future research should take a more comprehensive ecological approach that integrates a variety of environmental factors.

1. The ARIMA model reveals a declining population trend, with fluctuations linked to seasonal patterns.
2. Clustering results indicate higher sparrow densities in rural and semi-urban areas, while urban centres show a sharp decline.
3. Existing models highlight the importance of integrating additional ecological factors such as temperature, pollution, and vegetation cover for better accuracy.
4. Conservation measures like urban greening, installation of nesting spaces, and reduced pesticide use are essential for population recovery.
5. Long-term field studies and real-time monitoring should complement AI-driven predictions to develop effective conservation strategies.

Future Scope:

1. **Including Other Environmental Factors:** To improve forecast accuracy, future research should include elements like food availability, noise levels, temperature fluctuations, and air pollution.
2. **Improving Model Accuracy:** By integrating ecological simulations with AI-driven methodologies, it is possible to gain a better knowledge of sparrow population dynamics and how they react to environmental changes.
3. **Field-Based Validation:** Verifying forecasts and improving conservation tactics will be made easier by combining AI models with long-term monitoring and on-the-ground surveys.
4. **Study Region Expansion:** By extending studies outside of Maharashtra to several different regions, a more thorough grasp of sparrow population patterns in various habitats will be possible.
5. **Creation of Conservation Plans:** To support sparrow populations in diminishing locations, the results of AI-driven analysis can direct community-driven conservation initiatives, regulatory changes, and urban planning adjustments.

6. **Incorporation of Climate Change Analysis:** Analyzing how climate change is affecting bird populations over the long run might aid in forecasting hazards and creating mitigation strategies.

7. **Public Awareness and Citizen Science Initiatives:** Real-time sparrow population monitoring can be facilitated by involving communities through birding programs and crowdsourced data gathering.

8. **Cooperation with Conservation Organizations:** Large-scale conservation initiatives can be facilitated by forming partnerships with ecological research institutes and wildlife protection organizations.

9. **Policy Recommendations for Urban Planning:** Making use of research findings to support the incorporation of green spaces and artificial nesting sites in urban areas, among other bird-friendly city designs.

References/Citations:

1. Sharma, P., Binner, M., & Indian Council of Agricultural Research. (2020). *The decline of population of house sparrow in India*. International Journal of Agricultural Science, Volume 5.

2. Shansaz, U. H., Jr., Fazili, M. F., Wildlife Research Lab., & Postgraduate Department of Zoology, University of Kashmir. (2022). *Distribution and conservation issues of House Sparrow: A review*. International Journal of Advanced Scientific Research and Management, 7(9).

3. Jhajhria, A. & JEZS. (2020). *A review on the need for conserving the house sparrow in India*. Journal of Entomology and Zoology Studies, 8(2), 1157–1159.

4. Paul, M. R. & Consortium & Training Academy for Biosciences (CTAB). (2015). *A Review of House Sparrow Population Decline in India*. Asia Pacific Journal of Research, 38–40.

5. Modak, B. (2015). *Impact of Urbanization on House Sparrow Distribution: A Case Study from Greater Kolkata, India*. Proceedings of the Zoological Society, 70.

6. Balmori, A. & Hallberg, Ö. (2007). *The Urban Decline of the House Sparrow (Passer domesticus): A Possible Link with Electromagnetic Radiation*. Electromagnetic Biology and Medicine, 26(2), 141-151.

7. Biswal, D. (2021). *Urban Growth and the Population of House Sparrows and House Crows: Is the Urban Ecosystem Undergoing a Radical Change?* 42(1), 17-35.

8. Ali, S. and Ripley, R. (1987). Compact handbook of the birds of India and Pakistan. Oxford University press, Delhi, pp: 296-314.

9. Birdlife International (2013), “Passer domesticus," IUCN Red List of threatened species version 2013.2.

10. Dandapat, A., Banerjee, D., & Chakraborty, D. (2010). The case of the Disappearing House Sparrow (Passer domesticus indicus). Veterinary World, 3(2), 97–100. https://www.cabdirect.org/abstracts/20103054772.html