International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 13 Issue 3 ǁ March 2025 ǁ PP. 59-63

Dynamics of Human Migration Using Machine Learning Algorithm

G.Anto Abisha, S.Vibil, Dr. S. Arul Jothi, M.Sc., M.Phil.,Ph.D. NET.
Department of ComputerScience, Madurai Kamaraj University, Fatima College, Madurai,India.

***ABSTRACT**:*
Human migration is crucial, and everyone has a reason for migrating—some for education, some for family, and others for different purposes. The problem arises when people do not return to their native places, as migration can lead to overcrowded cities and the decline of rural areas. This study analyzes human migration patterns to predict whether a person who has migrated from their native place will return, using people's responses as a key factor in the analysis.

Date of Submission: 09-03-2025 Date of acceptance: 23-03-2025

I. INTRODUCTION:

Migration is happening every second in every country. However, it may cause overcrowding in cities and depopulation in rural areas. Predicting whether migrants will return, can help the government to plan infrastructure, housing, and public services in both urban and rural areas. It can also create job opportunities in native places to encourage return migration, maintain cultural and social stability in regions affected by migration, and assist in developing business and market strategies. A "Dynamics of Human Migration" utilizing machine learning with logistic regression aims to predict whether a person will migrate from native to another city or another country based on a set of personal and socio-economic factors. This algorithm is used for predicting binary outcomes (return to native/not return) by calculating the probability of migration based on the input features. Machine learning, particularly logistic regression, is a powerful tool for predicting migration based on the data. Several studies have examined human migration patterns using traditional statistical approaches. Recent advancements in artificial intelligence (AI) and machine learning (ML) have enabled more precise predictions of migration behaviors. Techniques such as logistic regression, decision trees, and neural networks have been applied to various social science problems, including migration prediction. This research extends these methodologies by utilizing logistic regression to classify returnees and non-returnees based on socio-economic factors.

II. DATA COLLECTION:

The data for this study was collected using Google Forms. Migrants were requested to complete an online survey to help predict migration patterns. The questionnaire included the following questions:

- What is your gender?
- What type of migration have you experienced?
- What was the primary reason for your migration?
- If you migrated for work, what industry are you in?
- Did you migrate alone or with family?
- How satisfied are you with your migration decision?
- Do you plan to migrate again in the future?
- If you migrate again, what would be your reason?

III. DATA PREPROCESSING:

To ensure data quality, the following preprocessing steps were performed:

3.1 Handling Missing data:

Rows with missing values in the target variable (**"will return to native"**) were removed.

```
df = df.dropna(subset=[target])
```

3.2 Encode Categorical variables:

Uses LabelEncoder() to convert categorical responses (e.g., "Male", "Female") into numerical values (0 ,1).

```
encoded_df = df.copy()
label_encoders = {}
for col in list(options_dict.keys()) + [target]:
    le = LabelEncoder()
    encoded_df[col] = le.fit_transform(df[col].astype(str))
    label_encoders[col] = le
```

3.3 Handling Class Imbalance using SMOTE (Oversampling)

SMOTE() is used to handle class imbalance in the dataset by generating synthetic samples for the Minority class.

```
smote = SMOTE()
X_train, y_train = smote.fit_resample(X_train, y_train)
```

3.4 Model Selection

This study employs **logistic regression**, a supervised learning algorithm used for binary classification problems. The model calculates the probability of migration return using the following logistic function:

$\mathbf{P(Y=1) = 1/1+e^{-(wX+b)}}$, where:

w represents the weights assigned to each feature,

X represents input variables,

b is the bias term.

IV. ANALYSIS:

The analysis of this study is based on the user input using logistic regression to predict the output. Calculate the prediction convert user input to numerical values then train the model and calculate its weights & bias.
The analysis of final prediction is:

1) Probability=0.525
2) P(y=1)>0.5, the model predict:

The person will return to their native place.

4.1 How prediction works

4.1.1 User Input to Numerical Values: Responses are encoded into numerical values**.**

Feature	User Choice	Encoded Value (X)
Gender	Male (1)	0
Migration Type	Temporary (1)	0
Reason for Migration	Education (2)	1
Industry	IT/Technology (1)	0
Alone/With Family	Alone (1)	0
Satisfaction	Very Satisfied (1)	0
Future Migration Plan	Yes, to another city (1)	0
Future Reason	Job Opportunities (1)	0

4.1.2 Model Weights & Bias: Model computes feature importance using statistical optimization.

Feature	Weight (w)
Gender	0.2
Migration Type	-0.3
Reason for Migration	0.5
Industry	-0.1
Alone/With Family	0.4
Satisfaction	-0.2
Future Migration Plan	-0.3
Future Reason	0.1
Bias (b)	-0.4

4.1.3 Probability Calculation:

The sigmoid function determines the probability of return migration

$= (0\times0.2)+(0\times-0.3)+(1\times0.5)+(0\times-0.1)+(0\times0.4)+(0\times-0.2)+(0\times-0.3)+(0\times0.1)+(-0.4)$

$= (0)+(0)+(0.5)+(0)+(0)+(0)+(0)+(0)+(-0.4)$

$= 0.1$

Sigmoid Function

$P(Y=1) = 1/ (1+e-0.11)$

$P(Y=1) = 1/ (1+0.90481)$

$= 1/ 1.90481$

≈ 0.525

4.1.4 Final Prediction

Probability = 0.525 (52.5%)

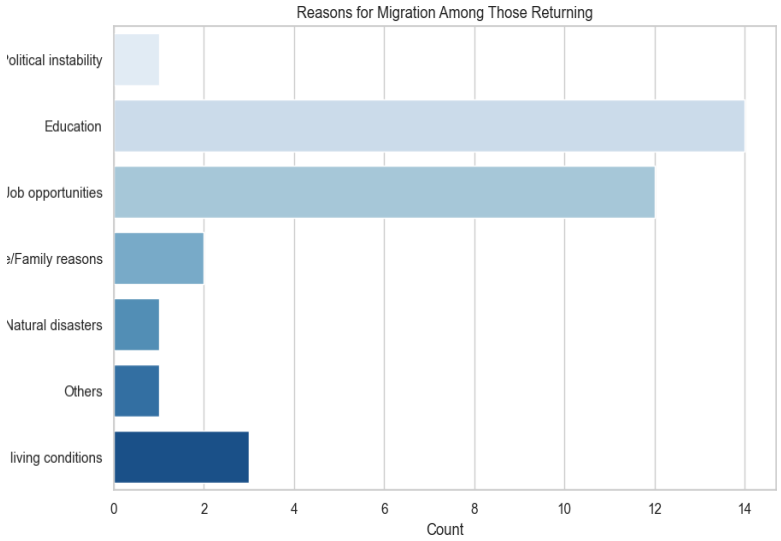
Since $P(Y=1)>0.5$

4.2. Analysis of dataset using Graph

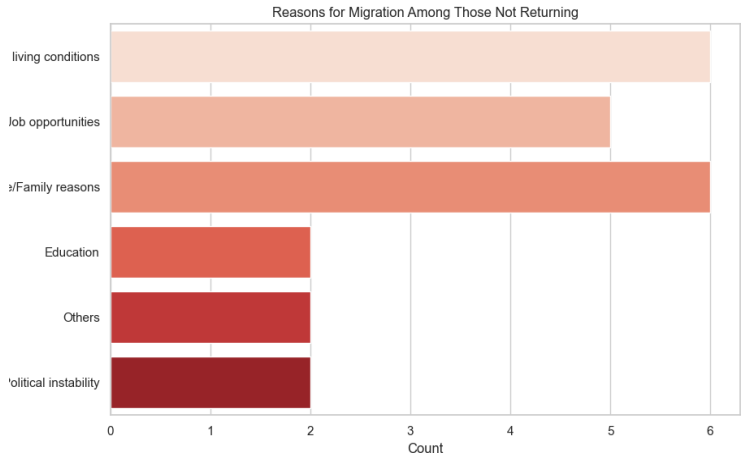
4.2.1 Distribution of Returnees

The model achieved an accuracy of 95% in classifying returnees and non-returnees. Job opportunities -(45%) and family reasons (30%) were the most significant factors influencing migration return.

4.2.2 Major Reasons for Migration among Returnees



4.2.3 Major Reasons for Migration among Non-Returnees



V. OUTPUT:

Output of the model is to predict a person is likely to return or not return to their native. The user give input to the questions and the system provides an automated prediction, if the prediction is greater than 0.5 then the person will return to their native or else the person will not return to their native, then the dataset is updated with new user entries.

```
Enter your choice (number): 3
Prediction: The person is likely to return to native.
✅ Dataset updated successfully in 'YES-dataset-migration.xlsx' with original user inputs.
PS D:\vibil\journal papaer>
```

VI. CONCLUSION:

This study of machine learning in predicting dynamics of human migration. Logistics regression proves to be an efficient technique for classification problems involving categorical data. The analysis reveals that personal and economic factors play a crucial role in migration decisions. While job availability is the dominant factor, family influence also significantly impacts return migration. By leveraging machine learning, this study offers a data-driven approach to understanding migration trends, which can be used for policy formulation, urban planning, and economic strategy development.

REFERNCES

[1]. Introduction to Machine Learning with Python: A Guide for Data Scientists by Andreas C. Müller and Sarah Guido
[2]. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow by Aurélien Géron
[3]. AI and Machine Learning For Coders: A Programmer's Guide to Artificial Intelligence by Laurence Moroney
[4]. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, by Sergey Ioffe, Christian Szegedy (2015) ICML. (cited 946 times, HIC: 56 , CV: 0).
[5]. Stock Market Prediction with High Accuracy using Machine Learning Techniques by Malti Bansal, Apoorva Goyal, Apoorva Choudhary.
[6]. Machine learning techniques and data for stock market forecasting: A literature review by Mahinda Mailagaha Kumbure [a], Christoph Lohrmann [a], Pasi Luukka [a], Jari Porras