Journal of Economics, Finance and Accounting Studies
ISSN: 2709-0809
DOI: 10.32996/jefas
Journal Homepage: www.al-kindipublisher.com/index.php/jefas

JEFAS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Artificial Intelligence in Socioeconomic Research: Identifying Key Drivers of Unemployment Inequality in the U.S

MD Abdul Fahim Zeeshan[1], Md Sumsuzoha[2], Faiaz Rahat Chowdhury[3] ✉ Md Rashed Buiya[4], MD Rashed Mohaimin[5], Laxmi Pant[6] and Reza E Rabbi Shawon[7]

[1]*Master of Arts in Strategic Communication, Gannon University, Erie, PA, USA*
[2]*Master of Science in Business Analytics, Trine University*
[3567]*MBA Business Analytics, Gannon University, Erie, PA, USA*
[4]*Master of Science in Cyber Security, California State University, Dominguez Hills*

Corresponding Author: Faiaz Rahat Chowdhury, **E-mail**: faiazrahatchowdhury@gmail.com

| ABSTRACT

Unemployment inequality remains one of the most vexing socioeconomic quagmires confronting the United States. This research project aimed to pinpoint how AI can be applied in the enumeration of key drivers of unemployment inequality in the United States and set a framework for further research and policy development. In this study, the researcher has drawn a massive volume dataset from the Economic Policy Institute's State of Working America Data Library, along with research performed by the Federal Reserve Bank of St. Louis. The unemployment incidents data was classified in terms of age, education level, gender, race, and other demographic factors. Subsequently, the analyst employed Linear Regression from the Scikit-learn library. Overall performance evaluation showcased that linear regression performed excellently with the least error in MSE and RMSE and, hence, was the best in terms of accurately predicting unemployment indicators. Accurate prediction of the unemployment rate using the proposed linear regression model can help the U.S. government proactively warn against economic downturns by deploying the. Besides, by executing the Linear Regression, government officials can influence favorable policies through tax incentives or labor laws. Evidently, the linear regression framework is a powerful AI tool that can help bring huge enhancements to unemployment inequality research and policy development in the future. This model not only provides a quantification of the relationships but allows for the making of predictions, thus making it useful for evaluating the possible results of different policy scenarios. Furthermore, the Linear Regression framework can also be used in the assessment of the effectiveness of pre-existing policies aimed at reducing unemployment.

| KEYWORDS

Unemployment Inequality; Socioeconomic issue; Artificial Intelligence; Linear Regression; Light GBM; XG-Boost; Supporting Vector Machines; Random Forest.

| ARTICLE INFORMATION

ACCEPTED: 02 September 2024 **PUBLISHED:** 23 September 2024 **DOI:** 10.32996/jefas.2024.6.5.6

1. Introduction

As per Khalaf et al. (2023), unemployment inequality persists as one of the most nagging socioeconomic predicaments confronting the United States. Nevertheless, not all groups in the United States experience unemployment equally. There are considerable inequalities across different dimensions, such as race, gender, education level, and geography. Chkwuere (2023) contends that improving unemployment inequalities can go a long way toward engendering greater economic prosperity and social justice in this country. In that regard, Artificial intelligence can be used to support research studies to understand the complicated drivers

of unemployment inequality. This research paper, therefore, aims to pinpoint some of the main factors responsible for unequal unemployment results in the U.S., with a focus on how A.I. techniques can facilitate socioeconomic research in this domain.
Körtner & Bonoli (2023) asserts that traditional economic analyses frequently fall short in pinpointing the multifaceted nature of unemployment because of the limitations in data handling and processing capabilities. Nevertheless, the advent of AI creates options for diving deep into socioeconomic issues. The consolidation of Artificial Intelligence (AI) in socioeconomic research has enhanced how researchers assess complex social phenomena, specifically unemployment inequality. This research project intends to highlight how AI can be applied to enumerate significant drivers of unemployment inequality in the U.S. and to set a framework for further research and policy development.

1.1 Problem Statement

Skandalis et al. (2022) hold that unemployment inequality in the U.S. remains a pressing matter, with the overall unemployment rate estimated at 3.8% as of 2023, yet in some groups, it is much higher. For instance, the unemployment rate in the case of Black Americans was about 6.0%, whereas for Hispanic people, it equaled 4.5%. Besides, a disparity may be observed by attainments in education level, with persons who have no more than a high school diploma having 5.5% unemployment rates compared with 2.1% for persons with college graduation. Young et al. (202) argue that there are also regional differences that make this inequality even more severe: states like Nevada and West Virginia have unemployment above 5.0%, which partly can be related to the economic decline in key industries. These factors interact in a very complex way, and most of the conventional analytics methods fail to study this interaction comprehensively or never get into doing so. This paper, therefore, calls for an approach capable of leveraging Artificial Intelligence to identify the key drivers of unemployment inequality. This paper targets the harnessing of A.I. techniques in uncovering actionable insight that may inform targeted policy interventions to foster a more equitable labor market.

2. Literature Review of Related Works

2.1 Overview of Unemployment Inequality

Unemployment inequality in the U.S. revolves around the substantial disparities in unemployment rates among different demographic categories and geographic regions. Although the national unemployment rate is an instrumental indicator of economic health, it often masks particular problems of inequality experienced by some disenfranchised populations. For instance, according to the Bureau of Labor Statistics, unemployment rates among Blacks were recorded at 6.0% at the beginning of 2023, which is significantly higher compared to a national average of 3.8% (Masoud, 2024). Other variables, such as educational attainment and geographic location, were significant factors in explaining such a diverse landscape, whereby some groups are affected more by economic fluctuations than others. Comprehending these disparities can be very instrumental in formulating proper policies geared toward actualizing equity within the labor market (Yurtsever, 2023).

Zezulka& Genin (2024) indicate that several interrelated factors can be attributed to the unemployment inequality in the United States. A major cause relates to educational attainment; higher attainment of education lessens the chances of becoming unemployed. For instance, in 2023, U.S. citizens with a bachelor's degree had an unemployment rate of around 2.1%, while persons with only high school attendance had an unemployment rate of 5.5%. This educational gap not only affects job access but also influences the types of jobs available to various demographic groups. Furthermore, regional economic conditions provide yet another critical factor. Areas whose economies are heavy in industries that are prone to economic downturns, such as manufacturing in the Rust Belt, exhibit a higher unemployment rate compared to regions that diversify their economy, such as Silicon Valley. All these factors add up to a very complex problem that will require an approach covering both the symptoms and the causes of unemployment inequality.

2.2 Unemployment as a Socioeconomic Issue in the U.S.

Unemployment is not just an economic problem but a deep socioeconomic that affects the well-being of citizens, families, and society at large in the United States. The repercussions of unemployment transcend beyond the loss of income; it entails interrelated an array of complex social, psychological, and economic impacts that may exacerbate poverty and inequality. The unemployment rate of the country, through 2023, is around 3.8%; although appearing to be low, it masks the large gaps between different demographic groups. For example, Black Americans have unemployment rates of 6.0% and Hispanic individuals at approximately 4.5% (U.S. Department of Labor, 2024). This disparity illustrates structural inequalities within the labor market, with unemployment intertwined with fundamental factors such as race, ethnicity, age bracket, education, and geographical location. The causes of unemployment can only be tackled by strategies that also take into consideration the many impacts of unemployment in the U.S.

One of the most irritating socioeconomic aspects of unemployment is its impact on poverty. According to the *U.S. Department of Labor (2024),* in 2022, the poverty rate was 12.4%, which translates to more than 37 million Americans living in poverty. Unemployment is considered one of the strong driving forces for residing in poverty because then people, after being unemployed, cannot afford such basic needs as housing, food, and healthcare. The economic strain of unemployment may involve a host of negative outcomes, including increased reliance on social safety nets, such as food assistance programs and unemployment benefits, further straining public resources. More importantly, the psychological side of unemployment, including stress, anxiety, and even depression, can make it even more difficult for these citizens to find new sources of employment, creating a sort of vicious cycle. The longer people are out of work, the more they become exposed to long-term financial instability and social isolation, conditions which further exacerbate the socioeconomic problems involved.

Education has emerged as the determining factor in employment. The huge disparity in educational attainment contributes highly to unemployment inequality in the U.S. Research studies show that unemployment is around 5.5% among people not possessing a high school diploma. In comparison, this rate quiets down to 2.1% among those with a bachelor's degree (U.S. Department of Labor, 2024). This education gap has its strong roots within marginalized communities where the quality of education remains an issue. For example, the National Center for Education Statistics reports that students from low-income families are less likely to graduate high school and go on to college, thus narrowing their options for future employment. As discussed in the section on education, the relationship between education, unemployment, and income explains why investing in educational opportunities can play a key role in addressing inequality in unemployment (Trading Economic, 2024). Policy measures directed towards ensuring access to quality education and vocational training could work out in creating a society with equality by equipping individuals with the needed skills to compete in a rapidly changing job market successfully.

2.3 The Role of A.I. in Socioeconomic Research

Kuikka (2020) contends that artificial intelligence has revolutionized the face of socioeconomic research by elevating data analysis capabilities. Several traditional methods are being challenged by data volumes and complexity that exist today. A.I. methodologies, particularly machine learning algorithms, can process large data volumes efficiently towards the identification of patterns and trends that are difficult to perceive through conventional statistical methods alone. This discovery means that AI, for example, can analyze economic indicators, demographic information, and employment data at the same time. With that capability, it finds the connection among variables such as education, industry, and regional economic health that researchers would not be able to see. That said, using A.I. enables researchers to bring a more detailed understanding of socioeconomic issues and understand phenomena such as unemployment inequality better.

Another critical contribution of AI in socioeconomic research is Natural Language Processing. NLP enables researchers to get valuable insights from unstructured data sources such as social media posts, news articles, and policy documents in the U.S. By analyzing the sentiment, trends, and public discourse on the economy, the researchers can sense the societal attitudes concerning unemployment, job creation, and the effectiveness of policy (Körtner & Bonoli, 2023). For instance, NLP may be applied to assess the impact of economic policy on different strata through evaluations of public comments and conversations over the Internet. This qualitative input supplements quantitative output and puts together a comprehensive perspective on socioeconomic processes for more effective policy intervention.

2.4 Machine Learning Algorithms in Socioeconomic Research

Decision trees are considered among widely deployed machine learning algorithms in socioeconomic research because of their interpretability and ease of use. They work by splitting data into subsets based on values of features and thus provide researchers with a clear view of the decision-making process. For instance, Khalaf et al. (2023) explored determinants of unemployment rates in various states in the U.S. using decision trees. Results from this research indicated that educational attainment and industrial composition were the relevant drivers of disparities in unemployment. This model identified highly educated states as having considerably lower rates of unemployment in the U.S., which is important to show how decision trees can find and present key relationships in data.

In their empirical research, Gabrikova et al. (2022) applied random forests in their work as part of the prediction between urban and rural divisions on unemployment rates in the U.S. According to the findings, the reasons that constituted the main causes of variation in unemployment included employment opportunities available, regional economic reasons, and demographic variables. The random forests provided insight into the ranking of variable importance to the model's most critical factors and thus suggested the targeted policy intervention the authorities could effectively make.

In another study by Zezulka & Genin [2024], the SVM algorithm was applied to classify regions from the U.S. based on their unemployment rate and other socioeconomic indicators. Their analysis showed that automation and technological progress have furthered unemployment inequality in highly manufacturing-based regions. The SVM model indicated steep nonlinear relationships between the variables, proving the robust capability of the algorithm in handling complex socio-economic data.

Young et al. (2023) implemented Neural Networks to analyze changes in unemployment rates across U.S. demographic groups driven by various education policy measures. The model was able to make use of a large set of features related to historical unemployment, educational attainments, and regional economic indicators that could pick up patterns that traditional methods may not have detected. The findings showed significant alleviation of disparities in unemployment in the U.S., especially for minority communities, resulting from increased investments in education. In this way, neural networks proved to be effective in uncovering those relationships and complexities that were hidden and not directly observable in the data.

Furthermore, Masoud (2024) used GBM to determine the determinants of unemployment inequality in America during the COVID-19 pandemic. The results indicated that industry type, job security, and working-from-home opportunities were important predictors of unemployment disparities in the United States. GBM models resulted in high values of predictive accuracy and interpretability while modeling the impact of a pandemic on particular segments of the labor market. After all, this capability for condition change essentially underlines the relevance of GBM for real-time socioeconomic analysis.

3. Methodology

In this research project, the following set of software tools was utilized: The Pandas library, which was used for data manipulation and analysis; the Python programming language; the Scikit-learn library, used for machine learning algorithms and metrics for the study; and the LIME library, which is essential for explainable A.I. Python was selected due to its versatility, simplicity, a great variety of choices for machine learning libraries, and extensive capabilities in data analysis [Pro-AI-Robikul, 2024]. Contrary to LIME, increasing the interpretability of machine learning models allowed the analyst to understand how the model generated a particular prediction.

3.1 Dataset

In this study, the researcher collected a large volume of datasets from the Economic Policy Institute's State of Working America Data Library and research performed by the Federal Reserve Bank of St. Louis; it comprises unemployment data segmented by age, education level, gender, race, and other demographic factors [Pro-AI-Robikul, 2024]. The dataset was based on a wide range of parameters related to unemployment inequality data, which helped the researcher to analyze its trend and pattern rigorously.

3.2 Pre-Processing

Data preprocessing involved eliminating missing values and applying a variety of techniques to clean the data collected. In this study, the analyst deployed the Min Max-Scaler technique imported from the library sci-kit-learn [Pro-AI-Robikul, 2024]. The role of the Min Max-Scaler was to transform attributes by scaling each attribute into a specific range. This approximator scaled and transformed every attribute independently to ensure it fell into the targeted range, usually between zero and one, basing its operations on the training set.

3.3 Feature Engineering

As per Pro-AI-Rokibul [2024], feature selection is one of the most indispensable processes that build out the predictive algorithm since the selection of features comprises the most relevant attribute, yielding the most impact on the stock price label. This whole process targets tapping the best subset of features, which will help drive a reasonably accurate stock price forecast. According to the researchers, by carefully choosing the right set of attributes, the algorithm captured all the critical relationships and, at the same time, patterns that were in the data. Having finished the feature selection stage, the very next phase entailed fitting the algorithm using selected attributes.

3.4 Algorithms Deployed

3.4.1 Linear Regression

Linear regression is a form of supervised machine learning in which an ongoing output is normally forecasted for a given variable input or sometimes multiple variable inputs. It models the relationship between the dependent variable, otherwise known as the target, and the independent variables, also known as the features, by fitting a linear equation to the observed data. That is, one variable will provide a line, and several will provide a hyperplane that minimizes the sum of the squared differences between the predicted and actual values (Chukwuere, 2023). This model trains a coefficient estimate for every input feature, with the goal that it captures the linear relationship that underlines it. Linear regression has become a staple in many developers' toolkits because it's very simple to conceptualize and interpret, and, where appropriate, can be used as the basis for predictive analytics.

3.4.2 Random Forests

Random Forest is an ensemble learning method that is frequently used for classification and regression. It constructs many decision trees during the training process and combines their results, anticipating better accuracy with a reduced risk of overfitting. At each step, a random subset of data and features is used to form trees, which increases the diversity among trees (Young et al., 2022). The final prediction is conducted by averaging the results (regression) or taking a majority vote (classification). It reduces variance and enhances the generalizing capability of the model on complicated datasets, improving its robustness and performance when there is noise and missing values.

3.4.3 Support Vector Machines

Masoud (2024) contends that the Support Vector Machine is considered a supervised machine learning algorithm applied in both classification and regression analysis. This methodology was developed to find the best hyperplane to separate classes of data features in multidimensional space. It does this by maximizing the margin, which is the distance between the hyperplane and the closest data points from each class, called support vectors. Maximizing this margin, therefore, SVM enhances the capability of generalizing the model for unseen data. SVM is efficient in handling non-linearly separable data by employing a kernel function that projects the data into a higher dimension, hence allowing the algorithm to find a linear separation in that space.

3.4.4 XG-Boost

XG-Boost is considered among the most efficient and popular machine learning algorithms for solving regression and classification problems. It is an efficient library interfacing with graduate-boosted decision trees for production goals, aiming at speed and accuracy. XG-Boost adopts this sequential ensemble approach whereby each added tree in the sequence reduces the progressive errors that were made by the predecessors (Kuikka, 2020). It further introduces the regularization methods to ensure that there is no overfitting during training, especially in highly dimensional and very big data. Some of the reasons it finds such extensive use in competitive modeling situations are parallel processing, tree pruning, and handling missing values.

3.4.5 Light GBM

Khalaf et al. [2023] indicate that Light-GBM is a high-performance, distributed gradient-boosting framework designed mainly for efficiency and scalability. This algorithm builds the trees sequentially by and at every step-in decision tries to correct the errors of the treetop and then combines the output of many weak learners to form a strong predictive model. It proposes a novel leaf-wise growth strategy instead of traditional level-wise growth that is feasible in both speed and memory usage and handles large datasets and high-dimensional data efficiently. It supports both parallel learning and GPU learning, which makes it faster than other boosting algorithms, such as XG-Boost, for large-scale tasks. On account of the low computational cost and high accuracy regarding big data, Light-GBM is widely used in classification, regression, and ranking tasks.

3.5 Experimentation Results

In Python, the analyst implemented five algorithms from the Scikit-learn library, most notably linear regression, Support Vector Machines, XG-Boost, light-GBM, and Random Forest Regression. All of the algorithms were imposed at a random state equal to zero and default parameters for the experiment's reproducibility.

3.5 Importing Libraries

*Import **Important** Libraries*

```
import pandas as pd
import numpy as np
from sklearn.preprocessing import StandardScaler, LabelEncoder
from sklearn.impute import SimpleImputer
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline
```

Load The Dataset

```
df = pd.read_csv("unemployed_population_1978-12_to_2023-07.csv")
```

```
df
```

To get a quick overview of the dataset, understand the distribution of the data, and identify outliers or unusual patterns. The following code snippet was computed:

```
df.describe()
```

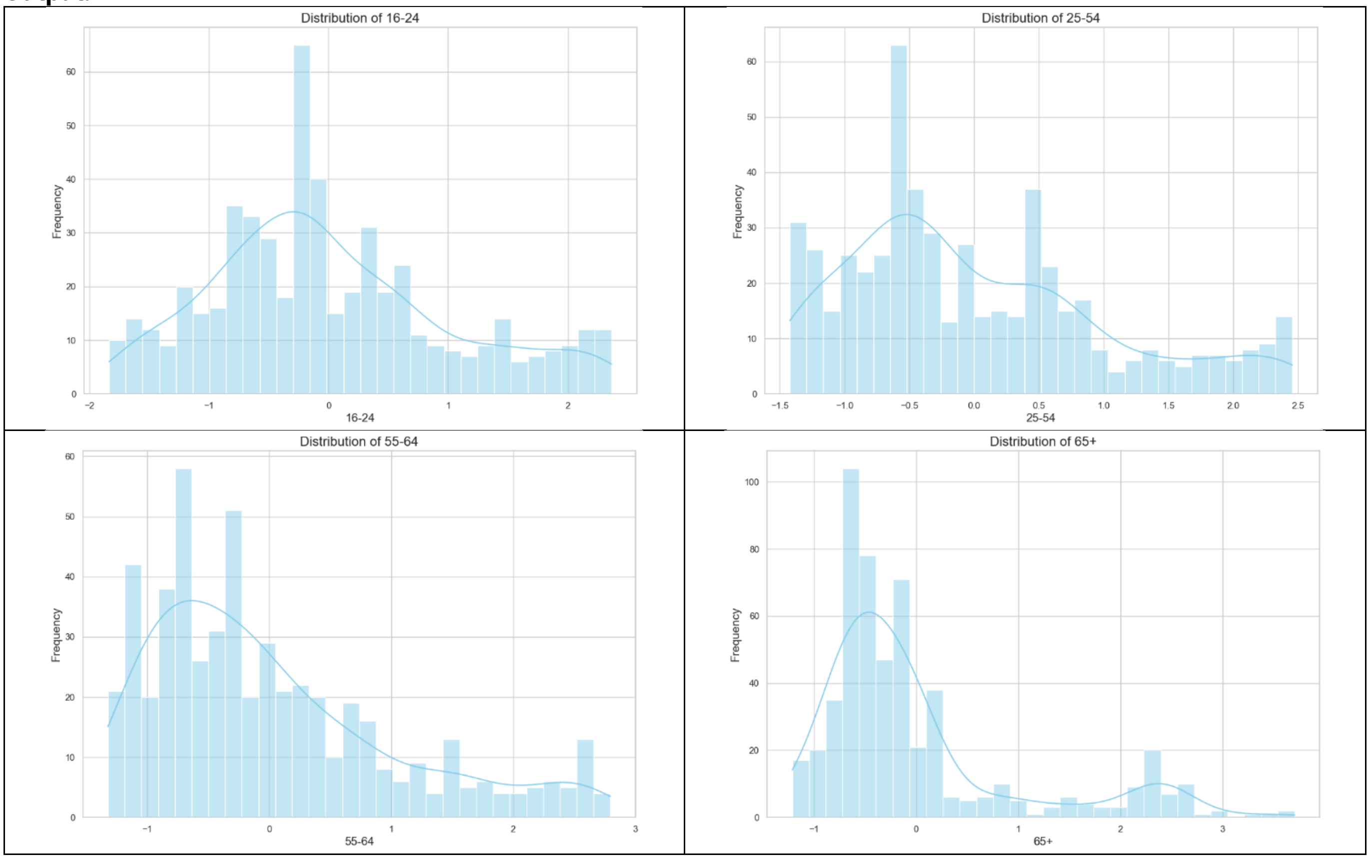
Output:

	all	16-24	25-54	55-64	65+	less_than_hs	high_school	some_college	bachelor's_degree
count	536.000000	536.000000	536.000000	536.000000	536.000000	536.000000	536.000000	536.000000	536.000000
mean	6.160261	12.558769	5.090672	3.972388	3.779291	12.366978	6.896455	5.350373	3.225000
std	1.633552	2.596647	1.470262	1.191348	1.141110	2.738656	1.918958	1.538527	0.923954
min	3.600000	7.800000	3.000000	2.400000	2.400000	7.000000	4.300000	3.100000	1.900000
25%	4.900000	10.800000	4.100000	3.100000	3.100000	10.400000	5.500000	4.200000	2.600000
50%	5.800000	12.100000	4.700000	3.700000	3.400000	11.900000	6.400000	5.000000	2.900000
75%	7.300000	14.000000	5.900000	4.600000	3.900000	13.900000	7.625000	6.100000	3.500000
max	10.300000	18.700000	8.700000	7.300000	8.000000	19.000000	12.200000	9.500000	6.400000

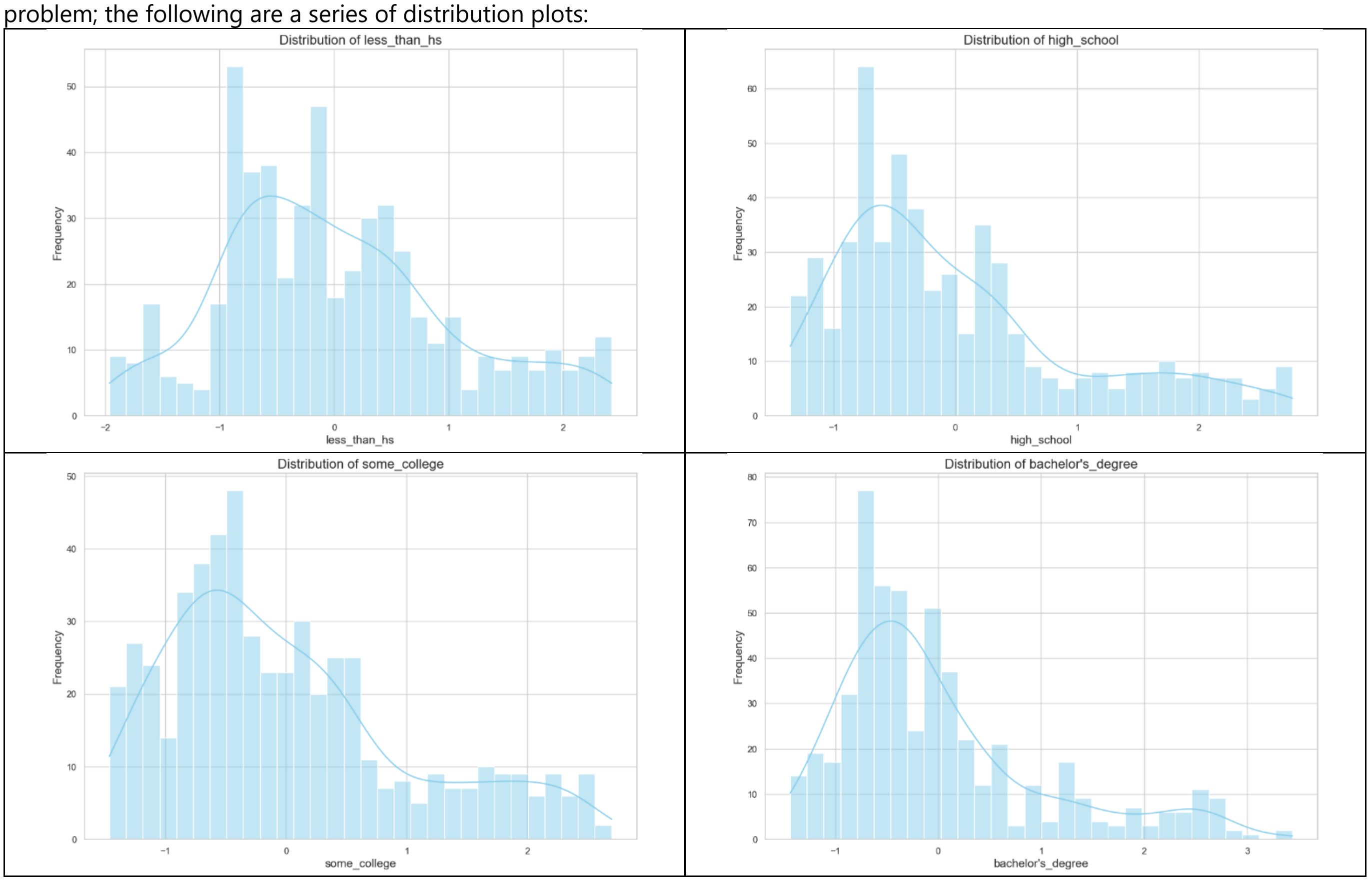
To generate a series of distribution plots for each numerical feature in the dataset. Exploratory data technique was applied to understand the distribution of all features, as showcased below:

```
import matplotlib.pyplot as plt
import seaborn as sns

# Set plot styles
sns.set(style="whitegrid")dtypes(include=[np.number]).columns
categorical_cols = df.select_dtypes(include=[object]).columns

# Target column (replace with the actual target column from your dataset)
target_column = 'all'

### 1. Distribution of Numerical Features ###
for col in numerical_cols:
    plt.figure()
    sns.histplot(df[col], kde=True
plt.rcParams["figure.figsize"] = (12, 8)

# List of Numerical and Categorical Columns
numerical_cols = df.select_
, bins=30, color='skyblue')
    plt.title(f'Distribution of {col}', fontsize=16)
    plt.xlabel(col, fontsize=14)
    plt.ylabel('Frequency', fontsize=14)
    plt.grid(True)
    plt.show()
```

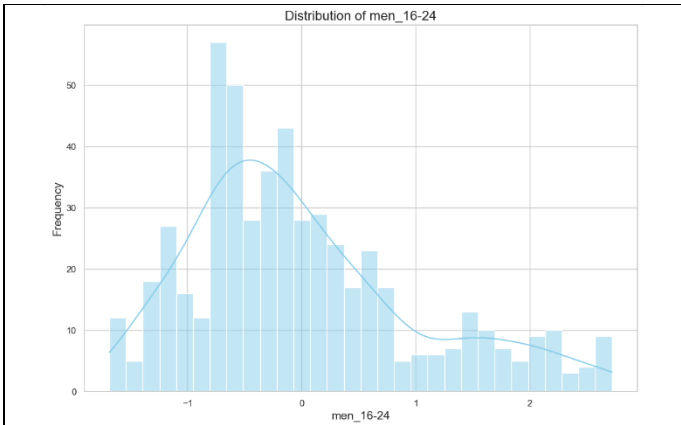
Output:

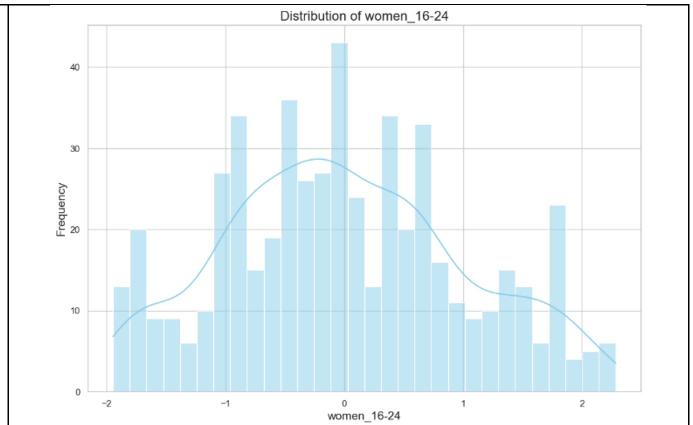


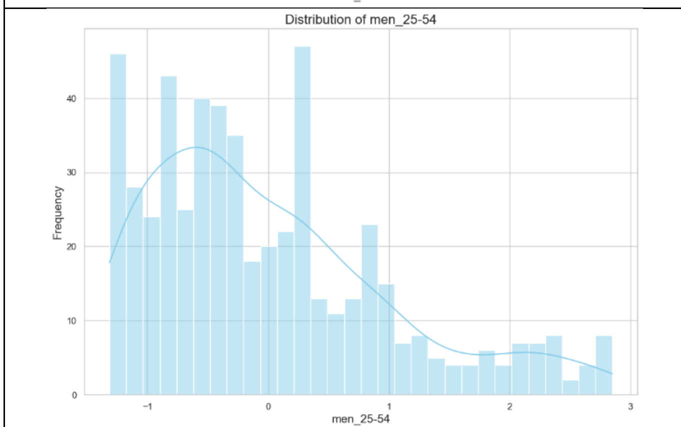
Apart from the age distribution, the analyst was also keen to assess the implications of education level on the unemployment problem; the following are a series of distribution plots:

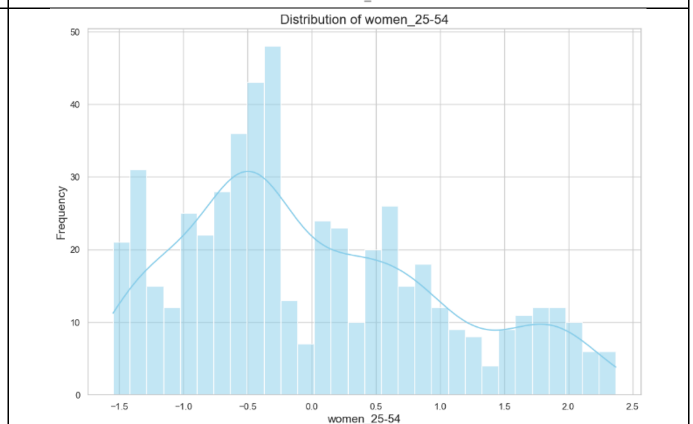


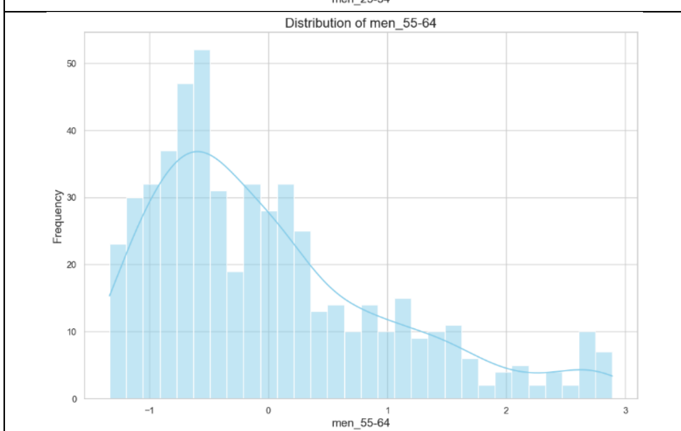
To explore the correlation between the age bracket of men and women and its implications on the level of unemployment, the following are some of the comparison distribution plots generated:

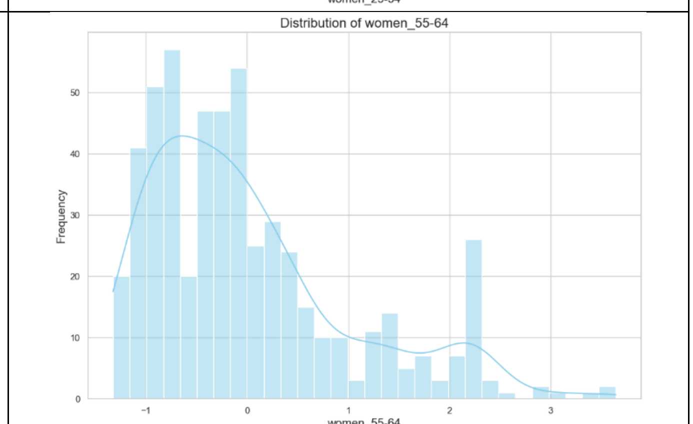


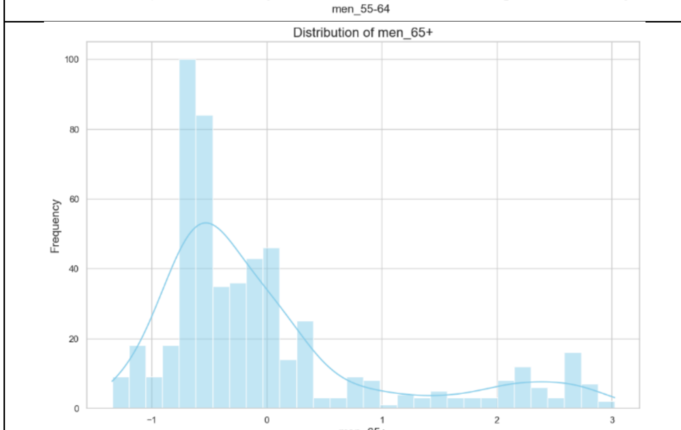


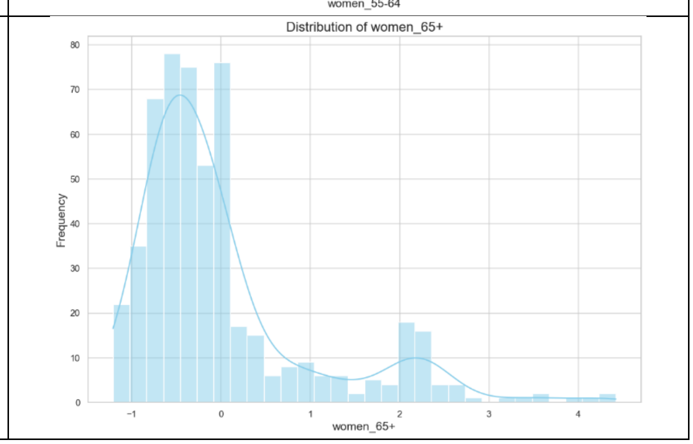




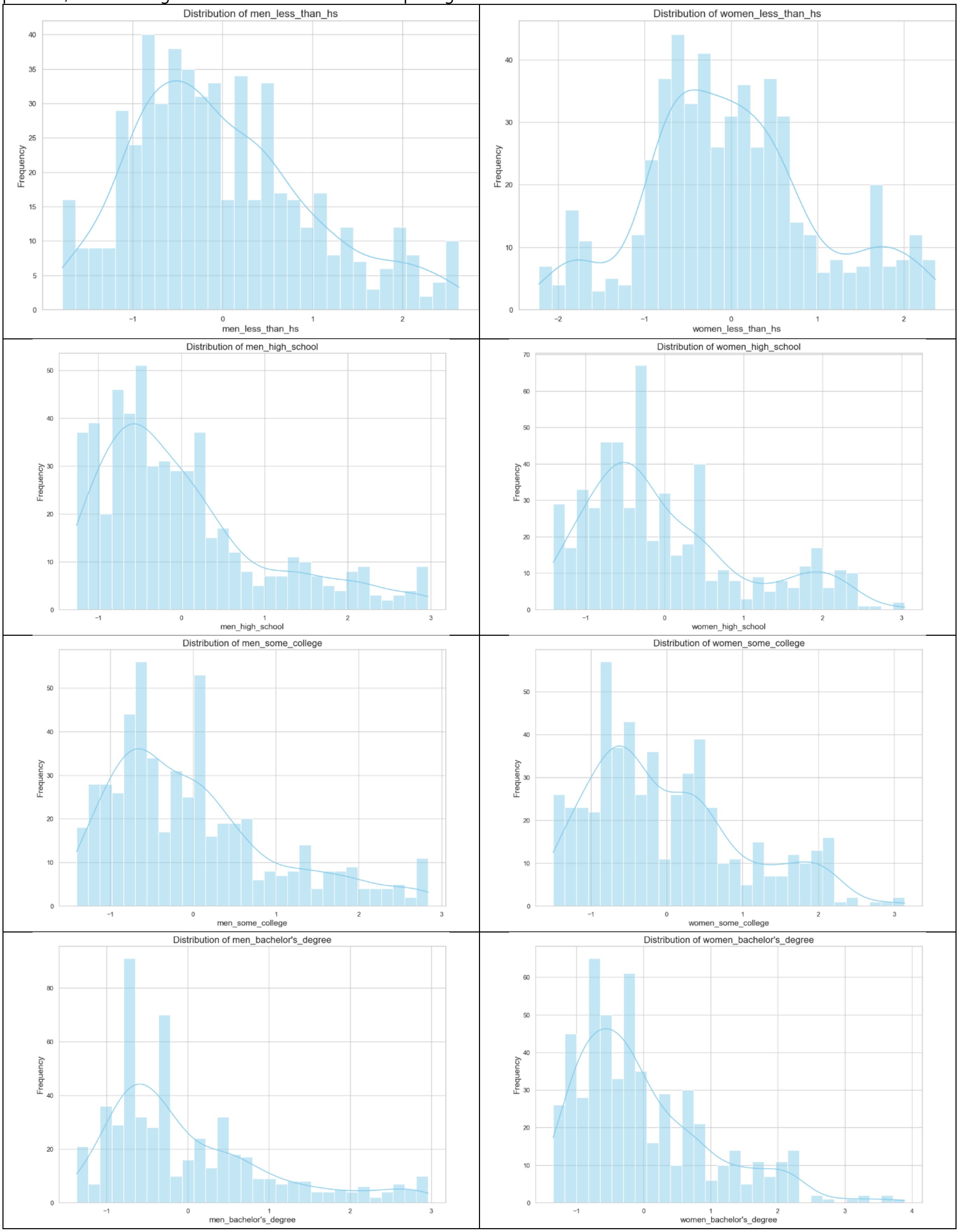


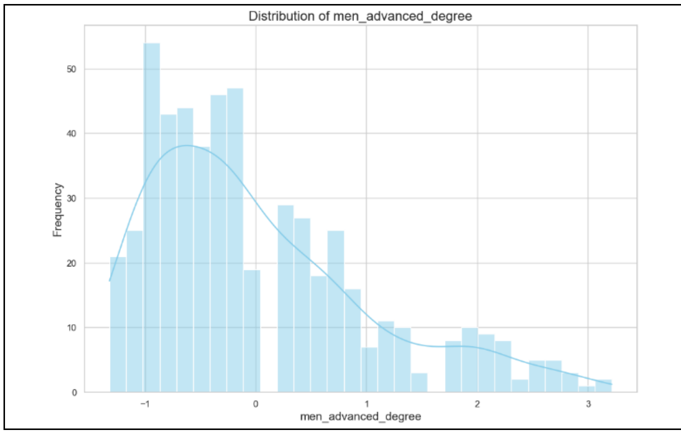


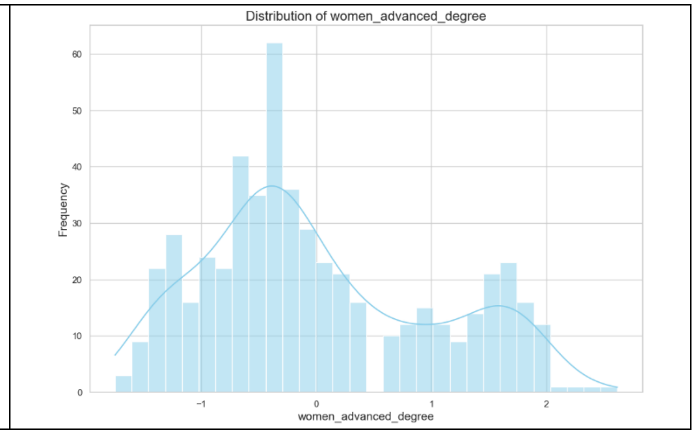




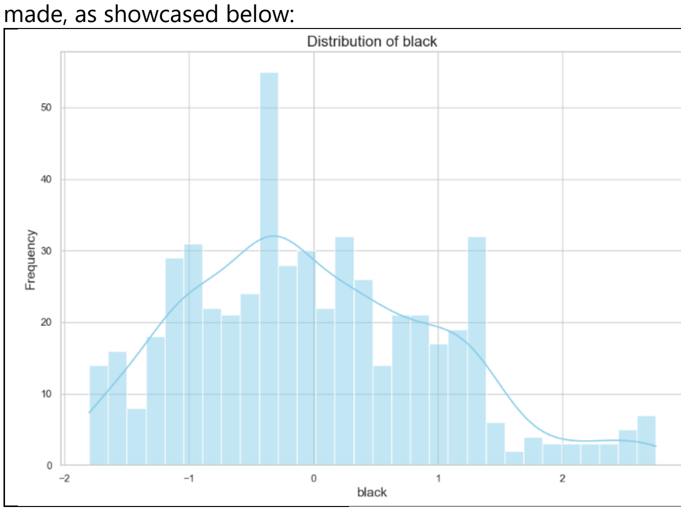
The analyst also explored the association between the impact of the educational level of men Versus Women on the unemployment problem; the following are some series of distribution plots generated:

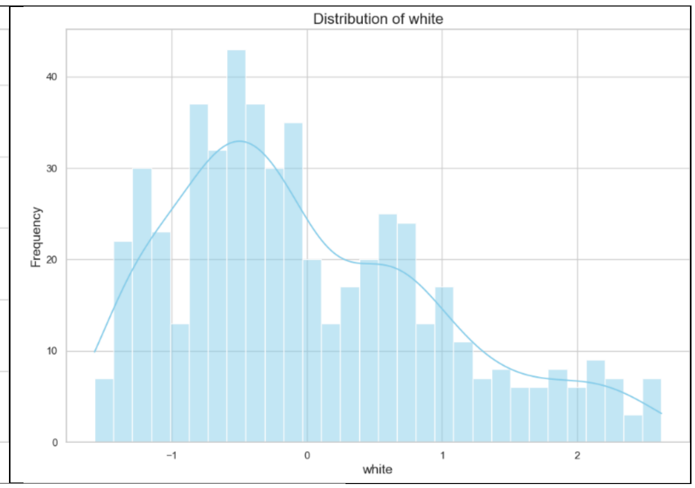


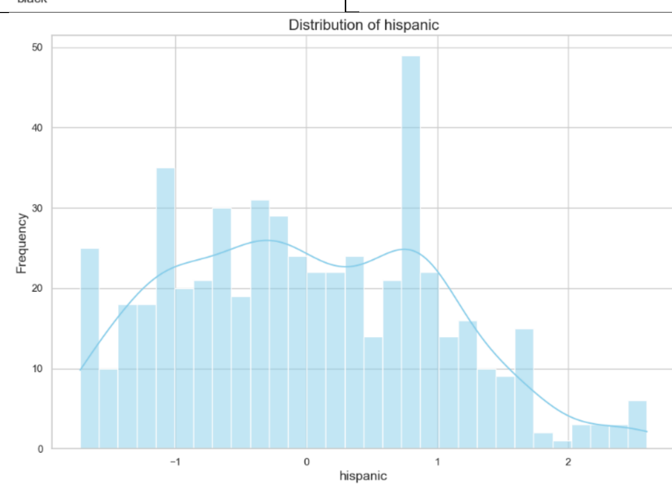




To obtain the overview correlation between race and the level of unemployment in the U.S., further exploratory analyses were made, as showcased below:







3.7 Model Performance Evaluation

- **Mean Squared Error [MSE]:** Denotes the Measure of the average squared difference between predicted and actual values. A lower MSE implies better performance.
- **Root Mean Squared [RMSE]:** Refers to the square root of MSE, portraying the error in the same units as the target variable. Lower RMSE is better.
- **R^2 Score:** This score indicates how well the model explains the variance in the target. An R^2 score closer to 1 is ideal, while negative scores indicate that the model is worse than a simple mean-based prediction.

Model	Mean Squared Error [MSE]	Root Mean Squared Error [RMSE]	R^2 Score
Random Forest	0.0021	0.0462	-0.0021
Support Vector Regressors	0.0026	0.0511	-0.0026
Linear Regression	0.0005	0.0219	-0.0005
XG-Boost	0.0019	0.0431	-0.0019
Light-GBM	0.0016	0.0399	-0.0016

Referring to the above table, linear regression performed well with the lowest error (MSE: 0.0005, RMSE: 0.0219), suggesting it performs best in terms of prediction accuracy. Light-GBM followed closely [MSE: 0.0016, RMSE: 0.0399], followed by XG-Boost [MSE: 0.0019, RMSE: 0.0431]. Random Forest and Support Vector Regressions had the highest error margins.

3.8 Framework for Future Research and Policy Development
A linear regression framework is a powerful AI tool that can greatly enhance unemployment inequality research and policy development in the future. It can establish the relationship between dependent and independent variables, identifying the contributing factors to unemployment disparity for appropriate interventions by researchers and policy developers.

3.9 Understanding the Framework
Linear regression attempts to model the relationship between a dependent variable, in this case, the unemployment rate, and one or more independent variables, such as education level, race, age bracket, geographic location, and economic conditions. The result from this model is a linear equation describing how changes within independent variables will affect unemployment. This not only provides a quantification of the relationships but allows for the making of predictions, thus making it useful for evaluating the possible results of different policy scenarios.

3.10 Identifying Key Driving Factors
One of the principal applications of linear regression in this context would be to ascertain the important determinants of unemployment inequality. For example, researchers can further research how education attainment and race may correlate with rates of unemployment across demographics and within them. Incorporating variables that include age, gender, and industry type can also show which of these factors disproportionately influence specific populations. This insight shall, therefore, be very significant in the formulation of policies that will address the actual causes of disparities in unemployment and ensure resources are allocated to serve the purpose.

3.11 Evaluating Policy Impact
A linear regression framework can also be used to assess the effectiveness of pre-existing policies aimed at reducing unemployment. In such a case, sections of data before and after the policy implementation would be considered, while changes in the rate of unemployment can be described as statistically significant or not. For instance, when introducing the job training program, linear regression can compare unemployment rates across regions where the program is implemented with those where it is not while controlling for other factors relevant to the study. Such evidence-based approaches go a long way in assisting policymakers with what works to ensure that evidence-based data drives decisions on future initiatives.

3.12 Benefits to the U.S. Economy

Early warnings of economic downturns: Accurate prediction of the unemployment rate using the proposed linear regression model can help the U.S. government proactively warn against economic downturns and provide sufficient time to adjust its strategies to avoid any financial loss in the future.
Risk Management: The government can effectively prepare for a high unemployment period through the linear regression models, where they can develop comprehensive risk management strategies such as getting finance or diversifying sources of revenue.
Policy Influence: With the right unemployment forecasts by deploying the Linear Regression, government officials can influence favorable policies through tax incentives or labor laws, amongst others, to cushion the effect of increased unemployment on their businesses.
Better Employee Retention Strategies: Unemployment forecasting will also enable the government to adopt employee retention strategies, such as providing competitive packages or benefits in cases where unemployment remains high and competition is greater.

4. Conclusion

Unemployment inequality persists as one of the most nagging socioeconomic predicaments confronting the United States. This research project intended to highlight how AI can be applied to enumerate significant drivers of unemployment inequality in the U.S. and to set a framework for further research and policy development. In this study, the researcher collected a large volume of data from the Economic Policy Institute's State of Working America Data Library and research performed by the Federal Reserve Bank of St. Louis; it comprises unemployment data segmented by age, education level, gender, race, and other demographic factors. In Python, the analyst implemented five algorithms from the Scikit-learn library, most notably, Linear Regression, Support Vector Machines, XG-Boost, Light-GBM, and Random Forest Regressor. Performance evaluation demonstrates that linear regression performed well with the lowest errors in MSE and RMSE, suggesting it performs best in terms of unemployment indicators prediction accuracy. Light-GBM was followed closely by XG-Boost, Random Forest, and Support Vector Regressors, which had the highest error margins.

References

[1] Chukwuere, J. E. (2023). I am exploring machine learning techniques to develop predictive models to address unemployment rates in the implementation of Industry 4.0. *Qeios*.

[2] De Troya, I. M., Chen, R., Moraes, L. O., Bajaj, P., Kupersmith, J., Ghani, R., ... & Zejnilovic, L. (2018). Predicting, explaining, and understanding the risk of long-term unemployment. In *NeurIPS Workshop on A.I. for Social Good*.

[3] Gabrikova, B., Svabova, L., & Kramarova, K. (2023). Machine learning ensemble modeling for predicting unemployment duration. *Applied Sciences*, *13*(18), 10146.

[4] HASAN, M. R. (2021). The Contribution of Foreign Direct Investment and Its Impact on Economic Growth and Inflation in Bangladesh. *PalArch's Journal of Archaeology of Egypt/Egyptology*, *18*(08), 4070-4088.

[5] Hasan, M. R., Shawon, R. E. R., Rahman, A., Al Mukaddim, A., Khan, M. A., Hider, M. A., & Zeeshan, M. A. F. (2024). Optimizing Sustainable Supply Chains: Integrating Environmental Concerns and Carbon Footprint Reduction through AI-Enhanced Decision-Making in the USA. *Journal of Economics, Finance and Accounting Studies*, *6*(4), 57-71.

[6] Karimian, H. R., Rouhanizadeh, B., Jafari, A., & Kermanshachi, S. (2019, June). A machine learning framework to identify employees at risk of wage inequality: U.S. Department of Transportation case study. In ASCE International Conference on Computing in Civil Engineering 2019 (26-34). Reston, VA: American Society of Civil Engineers.

[7] Khan, M. A., Debnath, P., Al Sayeed, A., Sumon, M. F. I., Rahman, A., Khan, M. T., & Pant, L. (2024). Explainable AI and Machine Learning Model for California House Price Predictions: Intelligent Model for Homebuyers and Policymakers. *Journal of Business and Management Studies*, *6*(5), 73-84.

[8] Khalaf, C., Michaud, G., & Jolley, G. J. (2023). Predicting declining and growing occupations using supervised machine learning. *Journal of Computational Social Science, 6*(2), 757-780.

[9] Körtner, J., & Bonoli, G. (2023). Predictive algorithms in the delivery of public employment services. In Handbook of Labour Market Policy in Advanced Democracies (387-398). Edward Elgar Publishing.

[10] Kuikka, S. (2020). Classification tree based algorithms in studying predictors for long-term unemployment in early adulthood: An exploratory analysis combining supervised machine learning and administrative register data.

[11] Masoud, N. (2024). Artificial intelligence and unemployment dynamics: an econometric analysis in high-income economies. *Technological Sustainability*.

[12] Pro-AI-Rokibul. (2024). *Unemployment-Inequality-in-U.S-Data-Analysis-And-Model/Model/Unemployment Inequality in US.ipynb at main · proAIrokibul/Unemployment-Inequality-in-U.S-Data-Analysis-And-Model*. GitHub. https://github.com/proAIrokibul/Unemployment-Inequality-in-U.S-Data-Analysis-And-Model/blob/main/Model/Unemployment%20Inequility%20in%20US.ipynb

[13] Skandalis, D., Marinescu, I., & Massenkoff, M. N. (2022). *Racial inequality in the U.S. unemployment insurance system* (No. w30252). National Bureau of Economic Research.

[14] Trading Economic. (2024). United States unemployment rate. https://tradingeconomics.com/united-states/unemployment-rate

[15] U.S. Department of Labor. (2024). *Employment and unemployment*. DOL. https://www.dol.gov/general/topic/statistics/employment

[16] Young, M. M., Himmelreich, J., Honcharov, D., & Soundarajan, S. (2022). Using artificial intelligence to identify administrative errors in unemployment insurance. *Government Information Quarterly*, *39*(4), 101758.

[17] Yurtsever, M. (2023). Unemployment rate forecasting: LSTM-GRU hybrid approach. Journal for Labour Market Research, 57(1), 18.

[18] Zezulka, S., & Genin, K. (2024, June). From the Fair Distribution of Predictions to the Fair Distribution of Social Goods: Evaluating the Impact of Fair Machine Learning on Long-Term Unemployment. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency* (1984-2006).