Assignment Report

Sadia Abdulhalim 60098686

Problem Definition:

This report investigates the feasibility of predicting a community's hardship index in Chicago using readily available socio-economic data. The hardship index serves as a pre-defined metric reflecting the overall socio-economic disadvantage faced by a community. The study explores the predictive capabilities of existing data, aiming to contribute to data-driven insights for targeted interventions, efficient resource allocation, informed policymaking, and proactive community development.

The research question addressed in this study is: "To what extent can readily available socio-economic data be leveraged to predict a community's hardship index in Chicago?" This inquiry is essential for deriving data-driven insights into socio-economic challenges, directing resources efficiently, informing policymaking, fostering community development, and enhancing planning and preparedness efforts.

Literature Review:

In [1], the study explores the symbiotic relationship between financial hardship and perceived stress, highlighting financial adversity as a precursor to negative perceptions of one's financial situation. Of particular significance is the identification of social support as a pivotal factor capable of intervening and reducing the levels of stress caused by financial hardship. By emphasizing the importance of considering both individual perceptions and external support networks, the study contributes to a nuanced understanding of how financial stress influences overall well-being.

In [2], the literature review on economic factors and suicide provides a comprehensive examination of the influence of economic downturns, such as recessions and unemployment, on suicidal behaviors. The findings underscore the need for targeted interventions, including unemployment benefits and minimum wage increases, to alleviate the mental health impacts of economic challenges. The review acknowledges the unique challenges posed by the COVID-19 pandemic, emphasizing the urgency of research to understand its potential impact on suicide rates and the effectiveness of existing interventions.

In [3], the reference shifts the focus to the impact of economic environments on families' financial stress and overall well-being. It critiques existing literature for often narrowing its scope to individual and family-level explanations, calling for a broader consideration of economic factors. The review advocates for the incorporation of theoretical frameworks that explicitly account for economic contexts, such as ecological systems theory. Additionally, it underscores the pivotal role of policy decisions in shaping economic environments and influencing their subsequent effects on families.

In summary, these references collectively highlight the multi-faceted nature of socio-economic challenges, shedding light on the importance of individual perceptions, external support networks, targeted interventions, and policy decisions in understanding and addressing financial hardship and its broader implications on community well-being.

Methodology:

For predicting a community's hardship index, we employed a systematic approach combining data acquisition, feature engineering, model development, and evaluation. Firstly, we extracted relevant socio-economic data from available datasets, focusing on features such as community area numbers, housing conditions, unemployment rates, education levels, and per capita income. Through careful feature engineering, we derived meaningful variables from the dataset, including the calculation of income per capita and average scores based on educational and safety metrics.

Next, we employed AutoML techniques, specifically H2O, to develop predictive models capable of forecasting a community's hardship index based on the engineered features. AutoML enabled automated hyperparameter optimization and model selection, leveraging its features to streamline the model development process. The trained models were then evaluated using various performance metrics, including mean squared error, root mean squared error, and mean absolute error.

Findings:

The analysis revealed several key findings regarding the predictive modeling of the community's hardship index. Firstly, H2O demonstrated remarkable efficiency, with a minimum training run time indicating the parallel processing capabilities of the algorithms used. The AutoML approach selected ensemble methods as the optimal approach, highlighting the importance of combining multiple algorithms for improved predictive performance.

In terms of preprocessing, the pipeline effectively handled both quantitative and categorical features, with missing values imputed using standard strategies. The final predictive model, based on Gradient Boosting Machine (GBM), achieved a mean squared error of approximately 0.00156 and a root mean squared error of approximately 0.0395 on the cross-validation data. These performance metrics indicate the model's effectiveness in accurately predicting the hardship index based on socio-economic factors.

Overall, the findings demonstrate the potential of predictive modeling techniques to forecast a community's hardship index using readily available socio-economic data. The developed model holds promise for informing strategic decision-making, resource allocation, and policy development aimed at addressing socio-economic disparities and fostering community development initiatives.

Conclusion:

This study suggests that leveraging readily available socio-economic data to predict a community's hardship index is feasible and holds potential for informing targeted interventions, resource allocation, policymaking, community development, and preparedness efforts. Further research is warranted to refine models and explore additional factors influencing hardship.

References

1. Mathieu S, Treloar A, Hawgood J, Ross V, Kõlves K. The Role of Unemployment, Financial Hardship, and Economic Recession on Suicidal Behaviors and Interventions to Mitigate Their Impact: A Review. Frontiers in Public Health. 2022;10:907052. DOI:

10.3389/fpubh.2022.907052. PMID: 35875017; P Park, N., Heo, W., Ruiz-Menjivar, J., & Grable, J. E. (2017).

2. Financial hardship, social support, and perceived stress. Journal of Financial Counseling and Planning, 28(2), 322–332. https://doi.org/10.1891/1052-3073.28.2.322MCID: PMC9298506.

3. Friedline, T., Chen, Z. & Morrow, S. Families' Financial Stress & Well-Being: The Importance of the Economy and Economic Environments. J Fam Econ Iss 42 (Suppl 1), 34–51 (2021). https://doi.org/10.1007/s10834-020-09694-9

SQL Code:

```sql
SELECT

  socio.COMMUNITY_AREA_NUMBER,

  socio.PERCENT_OF_HOUSING_CROWDED,

  socio.PERCENT_HOUSEHOLDS_BELOW_POVERTY,

  socio.PERCENT_AGED_16__UNEMPLOYED,

  socio.PERCENT_AGED_25__WITHOUT_HIGH_SCHOOL_DIPLOMA,

  socio.PERCENT_AGED_UNDER_18_OR_OVER_64,

  socio.PER_CAPITA_INCOME,

  schools.SAFETY_SCORE,

  schools.Environment_Score,

  schools.Instruction_Score,

      crimes.PRIMARY_TYPE,

  socio.HARDSHIP_INDEX


INTO dbo.FactTable

FROM

  dbo.chicago_socioeconomic_data AS socio

JOIN

  dbo.chicago_public_schools AS schools

ON

  socio.COMMUNITY_AREA_NUMBER = schools.COMMUNITY_AREA_NUMBER

JOIN dbo.chicago_crime AS crimes ON crimes.COMMUNITY_AREA_NUMBER =
socio.COMMUNITY_AREA_NUMBER;
```

```sql
-- Indexing the fact table
CREATE INDEX idx_CommunityAreaNumber ON dbo.FactTable(COMMUNITY_AREA_NUMBER);



ALTER TABLE dbo.FactTable
ADD INCOME_PER_CAPITA DECIMAL(18, 2)

UPDATE dbo.FactTable
SET INCOME_PER_CAPITA = PER_CAPITA_INCOME / PERCENT_HOUSEHOLDS_BELOW_POVERTY;



ALTER TABLE dbo.FactTable
ADD AVG_SCORE DECIMAL(18, 2);



UPDATE dbo.FactTable
SET AVG_SCORE = (SAFETY_SCORE + Environment_Score + Instruction_Score) / 3;

ALTER TABLE dbo.FactTable
DROP COLUMN Environment_Score;

ALTER TABLE dbo.FactTable
DROP COLUMN SAFETY_SCORE;

ALTER TABLE dbo.FactTable
DROP COLUMN Instruction_Score;
```

```
SELECT * FROM dbo.FactTable
```