A Thesis on Utilizing Machine Learning Models to Predict Material Hardship

By
Tirtha Das
22275001

A thesis submitted to the Department of Economics and Social Science in partial fulfillment of the requirements for the degree of Master of Science in Applied Economics(MSAE)

Department of Economics and Social Science
BRAC University

December 2023

Supervised By:

Taufiq Hasan Quaderia, Ph.D

Assistant Professor of Instruction,

Financial Planning Program at The University of Akron.

Submitted By:

Tirtha Das

22275001

Department of Economics and Social Sciences,

BRAC University

BRAC UNIVERSITY

Inspiring Excellence

DECLARATION

I, Tirtha Das, bearing Student ID 22275001, affirm that the content presented in this thesis is the result of my original research and has not been previously published or submitted to any institution. Any external literature, data, or work referenced in this thesis has been duly acknowledged and appropriately cited in the reference section.

Tirtha Das

Tirtha Das

Student ID-22275001

Student's Full Name & Signature:

APPROVAL

The thesis/project titled **“A Thesis on Utilizing Machine Learning Models to Predict Material Hardship”** submitted by *Tirtha Das, BRAC University, Id: 22275001* of Fall, 2023 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Masters of Science in Applied Economics on 17th December, 2023.

Examining Committee:	
Supervisor: (Member)	December 28, 2023
	Dr. Taufiq Hasan Quadria Assistant Professor of Instruction Finance Planning Program Department of Finance, College of Business The University of Akron
Graduate Program Coordinator: (Member)	
	Dr. A S M Shakil Haider Assistant Professor Department of Economics and Social Sciences Brac University
External Expert Examiner: (Member)	Abdul Hannan Chowdhury
	Dr. Abdul Hannan Chowdhury Professor School of Business & Economics North South University
Departmental Head: (Chair)	M.W.R. Khan
	Dr. Wasiqur Rahman Khan Professor and Chairperson Department of Economics and Social Sciences Brac University

Abstract

This thesis addresses the concern of material hardship, defined by a paucity in resources necessary to fulfil fundamental needs, especially affecting children and families. Drawing insights from studies on the determinants of material hardship, including low income, unemployment, single parenthood, and financial literacy, the research employs a comprehensive methodology. It incorporates the integration of machine learning techniques to enhance the predictive capacity of identifying at-risk populations. The methodology advocates for a holistic approach, incorporating the transformative potential of machine learning techniques such as Logistic Regression, Non-Linear Support Vector Machine Model, Decision Tree etc. This paper highlights the transformative potential of machine learning in proficiently analyzing extensive datasets to recognize complex patterns. The positive correlation established between higher financial literacy, bill paying tendency, management of finances and improved economic outcomes stresses the potential impact of targeted financial education initiatives. Moreover, the research emphasizes the need for a proactive stance, advocating for the development of predictive models using historical evidence to anticipate and address material hardship in a timely manner. The inferences emphasize the necessity for targeted interventions and proactive measures, promoting social equity, resilience, and contributing to broader poverty reduction strategies.

Table of Contents

1. Introduction

Material hardship refers to the paucity of resources needed to meet fundamental needs, such as food, healthcare, housing, and transportation. Finding those who are at risk of experiencing material hardship is vital for developing effective involvement and support schemes. Children in families that encounter material hardship—unable to fulfill fundamental needs such as shelter, food, and medical care—face substantial challenges. Hardship has not only been linked to poor health, cognitive, and behavioral outcomes among children in the short term, but it may also contribute to the negative association between childhood poverty and long-term health and economic prospects (Ashiabi and O'Neal 2007; Chaudry and Wimer 2016; Gershoff et al. 2007; Ratcliffe 2015). Hardships in early childhood may be particularly deleterious given the rapid brain development that occurs during this period (Duncan, Ziol-Guest, and Kalil 2010; Shonkoff, Boyce, and McEwen 2009). Identifying individuals at risk of material hardship is key for targeted support, appropriate interventions, better outcomes, policy development, social equity, and cost savings. By identifying and supporting those in need, we can work towards a more inclusive and resilient society. Conventional methods for identifying at-risk populations often rely on self-reported data or manual screening processes, which can be time-consuming, subjective, and prone to error. However, the integration of machine learning techniques offers an opportunity to improve the accuracy and competency of finding individuals who are at risk of material hardship. "Machine Learning for Poverty Reduction: A Review of the Literature" by A. Jain and A. S. Mishra (2022) provides a comprehensive overview of how machine learning has the potential to transform poverty reduction by addressing the challenges of lack of data, complexity of poverty, and the need for targeted interventions. From the study, it has been found that utilizing machine learning can be an efficient tool in predicting poverty levels, recognizing people at risk of poverty, and monitoring the progress of poverty reduction programs This research will try to explore the factors that might influence material hardship and will try to build a predictive model on the survey database of the National Financial Well-Being Survey.

The paper by Beverly (2001) found that low income is one of the most important factors, but other factors such as unemployment and single parenthood also play a role in material hardship. The study by Hastings, Madrian, and Skimmyhorn (2013) explores the connection between financial literacy, financial education, and economic outcomes. Their discoveries show that individuals with higher levels of financial literacy tend to make better financial decisions, leading to improved economic outcomes. There are several factors that may influence material hardship, like low income, high debt, and unexpected expenses. However, it is difficult to forecast who is most at risk of material hardship. Machine learning is a commanding tool that can be used to classify people who are at risk of material hardship based on historical evidence. Machine learning algorithms can study large datasets of data to detect patterns that would be tough to find manually. This information can then be used to develop predictive models that can find individuals who are most likely to encounter material hardship.

2. Literature Review

Material hardship indicates the lack of resources needed to meet fundamental needs, such as food, healthcare, housing, and transportation. Finding those who are at risk of experiencing material hardship is vital for developing effective involvements and support schemes. In a paper from 2022 by Margaret M. C. Thomas, material hardship is examined as a direct indicator of deprivation in the United States, supplementing income poverty to give light on how deprivation affects wellbeing. The research examines five categories of material hardship encountered over a 15-year period, including difficulties with food, housing, healthcare, utilities, and bill-paying. It does this by examining data from the Fragile Families and Child Well-Being Study. Children in families that encounter material hardship—the incapacity to meet fundamental needs such as shelter, food, and medical care—face substantial challenges. The inability to access goods or services deemed necessary for respectable human functioning is the definition of material hardship in the 2012 paper "Measuring Poverty: The Official U.S. Measure and Material Hardship" by Gesemia Nelson. Contrary to income-based indicators like the official poverty line, the author claims that material hardship offers a more precise indicator of poverty. The study shows that material hardship affects more people than official poverty statistics suggest, and it extends to groups of people who are not typically thought of as poor in addition to traditionally recognized impoverished groups. Many magnitudes of hardship, including those related to food, health care, and bill-paying, have been considered in many earlier studies, but there is wide variation in whether those dimensions were treated as discrete hardship experiences. The study conducted by Colleen Heflin, John Sandberg, and Patrick Rafail, published in Social Problems in November 2009 observes the structural consistency of material hardship by comparing and testing five conceptual models. It finds that a multidimensional model, which separates material hardship into distinct dimensions such as health, food, bill paying, and housing, fits the data better than unidimensional approaches. In a study showed by Beverly (2008), titled "Measures of Material Hardship: Rationale and Recommendations," the author provides a comprehensive inspection of material hardship and its consequence in the background of social research and policy analysis. The paper highlights the importance of accurate measurement to comprehend the prevalence and significances of economic deprivation, guiding policymakers in formulating targeted interventions to lessen hardship and improve overall well-being. In their study, McKernan, Ratcliffe, and Braga (2021) examine the impact of the US safety net on material hardship over a span of two decades.

Their longitudinal analysis delves into the efficiency of various social safety net programs, such as TANF, SNAP, and housing assistance, in reducing material hardship among susceptible populations in the United States. Hardship has not only been linked to poor health, cognitive, and behavioral outcomes among children in the short term, but it may also contribute to the negative association between childhood poverty and long-term health and economic prospects. The connection between children's health status and material hardship is observed at in the study by Ashiabi and O'Neal (2007). Beyond income poverty, the research highlights the significant influence of material hardship on children's health outcomes. The writers of the paper "Poverty is Not Just an Indicator: The Relationship Between Income, Poverty, and Child Well-Being" by Chaudry and Wimer (2016) observe the complex association between money, poverty, and child well-being. The paper highlights the implication of considering poverty not merely as a statistical measure but as a critical determinant of children's welfare and future prospects. The patterns and predictors of material hardship among poor families with children were inspected in the study conducted by Katherine E. Marçal, which was published in December 2022.A distinction was made between three subtypes: "Housing Insecure," "Food Insecure," and "Cost-Burdened but Managing." The negative effects of maternal IPV exposure and depression on coping mechanisms highlight the significance of maternal safety and mental health in handling material hardship. Given the fast brain development that takes place at this time, early childhood hardships may be especially harmful (Duncan, Ziol-Guest, and Kalil 2010; Shonkoff, Boyce, and McEwen 2009).. Zilanawala and Pilkauskas (2012) observed material hardship as a predictor of children's socioemotional behavior and observed that hardship predicted more internalizing and externalizing behavior issues. The connotation between material hardship and child behavior is sturdier at age 5, and chronic aggregate hardship has a more profound influence on child behavior. "Material Hardship and the Physical Health of School-Aged Children in Low-Income Households" by Joan P. Yoo, Kristen S. Slack, and Jane L. Holl found that material hardship, and most powerfully food hardship, to be predictive of worse parent-reported health status among children.

The paper by Beverly (2001) found that that low income is one of the most important factors, but other factors such as unemployment, single parenthood also play a role in material hardship. The study by Hastings, Madrian, and Skimmyhorn (2013) explores the connection between financial literacy, financial education, and economic outcomes. Their discoveries show that individuals with higher levels of financial literacy tend to make better financial decisions, leading to improved economic outcomes. Gershoff, Aber, Raver, and Lennon (2007) noted in their study that only income may not accurately reflect the financial problems that families face and material hardship, which is the inability to pay for necessities. The paper titled "The relationship between income and material hardship" by James X. Sullivan, Lesley Turner, and Sheldon Danziger published in 2007 studies the association between income and material hardship among welfare recipients, highlighting that higher income is associated with reduced hardship, but temporary income changes have weak effects. It found that a 10 percent increase in average income over the 6-year study period is related with a 1.1 percentage point decrease in the likelihood of suffering material hardship, which amounts to a reduction of roughly 3.4 percent. There are several factors that may influence material hardship like low income, high debt, and unexpected expenses. However, it is difficult to forecast who is most at risk of material hardship. Machine learning is a commanding tool that can be used to classify people who are at risk of material hardship from historical evidence. Colleen M. Heflin's research, which was published in Social Problems in November 2017, examines how social positioning impacts patterns of material hardship. Utilizing data from the 2008 Survey of Income and Program Participation, this study observes long-term trends in material hardship. The results reveal both one-time and recurring hardships across domains. Hardships are consistently predicted by demographics like disability status, while other factors have different effects on domains and recurrence.

Identifying individuals at risk of material hardship is key for targeted support, appropriate interventions, better outcomes, policy development, social equity, and cost savings. By identifying and supporting those in need, we can work towards a more inclusive and resilient society. Conventional methods to identifying at-risk populations often rely on self-reported data or manual screening processes, which can be time-consuming, subjective, and prone to error. However, the integration of machine learning techniques offers an opportunity to improve the accuracy and competency of finding individuals who are at risk of material hardship. The benefits to social welfare that come from employing ML to enhance estimates of worker productivity are shown by Chalfin et al. (AER, 2015). The article "Machine Learning for Economics Research: When What and How?" by Ajit Desai (2023) mentions that the digitization of the economy has led to an explosion of economic data, posing challenges and opportunities for analysts. Machine learning (ML) has appeared as a powerful tool to handle huge and complex datasets, making it gradually popular in economics research. The use of ML in academic publications has remarkably risen in recent years. Susan Athey's chapter (2018) on machine learning (ML) and economics covers this topic. The article emphasizes how ML is especially beneficial for studying unconventional and unstructured data, capturing significant nonlinearity, and increasing prediction accuracy. Although the algorithmic nature of machine learning (ML) makes model selection and tweaking flexible, applying it to causal implication necessitates specialized algorithms and dependable confidence intervals. Economic theory is projected to undergo major changes as a result of the use of ML and new data sets, including new issues, methods, partnerships, and the implementation of new policies. This research will try to explore the factors that might influence Material Hardship and will try to build a predictive model on the survey database of National Financial Well-Being Survey.

Machine learning algorithms can inspect large datasets of data to notice patterns that would be troublesome to find manually. This information can then be used to develop predictive models that can pick out people who are most likely to encounter material hardship. In addition to identifying people who are at risk of material hardship, the predictive model could also be used to track the efficacy of interventions designed to reduce material hardship, and to raise awareness of the issue of material hardship in the United States. This could lead to increased public support for policies that aim to reduce material hardship.

3. Methodology

3.1. Data Set Description

The dataset chosen for this paper was one that had enough connection to the actual data that is probably going to be available, so I knew that my model design would be able to accommodate the kinds of citizen profile data that would be utilized in real-life implementation. The primary requirements for an acceptable dataset included having several other variables frequently found in government datasets, together with a qualifying variable to assess financial well-being.

The dataset selected is from a survey conducted by the **Consumer Financial Protection Bureau (CFPB)** published in September 2017. The dataset consists of 217 columns and 6395 rows.

3.2. Variable Description

3.2.1. Dependent Variable

Material Hardship (Sum of MATHARDHIP_1-6)

This thesis defines the dependent variable as 'Material hardship,' which comprises six essential questions intended to gauge the level of financial stress experienced by individuals or households. Respondents offer values between -1 and 3, with a score of 3 indicating 'Often' frequency for facing a particular type of hardship and 1 indicating 'Never' in frequency. Respondents who choose -1 are indicating a refusal to answer. The material hardship comprises of 6 questions. (Exhibit: 12)

I. Worried whether food would run out before got money to buy more

II. Couldn't afford a place to live

III. Any household member could not afford to see doctor or go to hospital

IV. Any household member stopped taking medication or took less due to costs

V. Food did not last and did not have money to get more

VI. Utilities shut off due to non-payment

The scoring system assigns values to each response, resulting in a cumulative score ranging from 3 to 18. A higher total score indicates a greater degree of material hardship, reflecting the severity of the financial challenges faced by the individual or household. Conversely, a lower score suggests a relatively lower level of financial strain. This dependent variable, with a minimum score of 3 and a maximum score of 18, provides a quantifiable way to assess the degree of financial hardship experienced by participants.

3.2.2. Independent Variable

I will try to incorporate financial knowledge and skill-based question from the survey. Also including variables that will influence the predictions to be more accurate.

Handling major unexpected expense by Individuals (FWB1_1)

FWB1_1 is a key statistical variable for predicting individuals with risk of facing material hardship. The responses stored range from -4 to 5. The responses are in ascending order where 5 indicates the severity of handling a major expense. -4 and -1 are responses not written in database and the respondents refused to answer the question. It reflects participant's ability of handling a major expense which is crucial to predict an individual's chance of facing a hardship in life.

Securing financial future by an Individual (FWB1_2)

FWB1_2 is a key determinant for predicting whether an individual might face material hardship or not. The responses stored range from -4 to 5. The responses are in ascending order where 5 indicates the capability of securing a financial future. -4 and -1 are responses not written in database and the respondents refused to answer the question. The responses indicate a participant’s ability of securing a financial future for themselves. This will assist a machine learning model to predict better.

Giving a gift...would put a strain on an individual’s finances for the month (FWB2_1)

One important factor in determining whether a person may experience material hardship is FWB2_1. The cached replies are in the range of -4 to 5. The answers are listed in ascending order, with 5 denoting a person’s positive response to the question. The responses -1 and -4 are not recorded in the database, and the respondents declined to respond to the inquiry. The answers show a person’s take on giving gift and how it would influence his/her finances for a month.

Respondent have money left over at the end of the month (FWB2_2)

Another variable used for the analysis is FWB2_2. The responses stored range from -4 to 5. The responses are in ascending order where 5 indicates the tendency of having money at the end of a month. The responses -1 and -4 are not recorded in the database, and the respondents declined to respond to the inquiry. This variable gives an indication of spending and savings characteristics of an individual. This can be a key factor of an individual’s possibility of facing hardship.

Financial skill-based Questions (FS1_1 to FS2_3)

Financial skill-based questions are included to make a better model to predict an individual’s risk of facing hardship. This variable represents the degree of competence a person believes they possess in different areas of money management. The rating includes the following abilities: making routine and innovative financial decisions, finding pertinent counsel, recognizing good investments, exercising discipline in spending, and saving money proactively. Self-awareness of information gaps and willingness to ask for help when needed are also included in the variable. This variable reflects both areas of potential improvement and strengths in financial competency, providing a comprehensive picture of the subject. These variables have a scale from -1 to 5. -1 response does not provide any insight as it represents no response.

Respondent’s Financial Commitment Behaviour (ACT1_1)

An emphasis on a individual's behaviour with regard to financial obligations to others is proposed by the variable "ACT1_1". If someone were to answer "Completely" to this statement, it would mean they have a history of keeping their financial commitments and pledges. It displays a feeling of accountability and dependability in honouring monetary promises made to other parties. The scale ranged high at 5 indicate that an individual follows through on their financial commitments to others.

Commitment to financial goals by Individuals (ACT1_2)

The variable "ACT1_2" pertains to an individual's commitment to personal financial goals. A positive response to this statement indicates that the person is diligent in following through on the financial objectives they set for themselves. It reflects a proactive and disciplined approach to achieving personal financial milestones. This variable captures an individual's ability to adhere to their own financial plans, showcasing a sense of self-discipline and determination in working towards their financial aspirations. The scale ranged high at 5 indicate that an individual follows through on their own financial goals.

Financial Goal Vision by Respondent? (FINGOALS)

The variable "FINGOALS" addresses whether an individual currently has or had a recent financial goal. It serves as an inquiry into the presence or absence of specific financial objectives within a given timeframe. A response of 1 indicates that the person has set or had a recent financial goal, while a response of 0 suggests the absence of such a goal during the specified period. This variable provides insight into an individual's engagement with financial planning and goal-setting activities and thus will assist to predict material hardship.

If respondents consult their budget to see how much money they have left (PROPPLAN_1)

The "PROPPLAN_1" variable aims to assess individual differences in financial planning behaviours. This item emphases on the practice of regularly checking one's budget to monitor the remaining funds. A positive response which is 5 in dataset shows a proactive approach to financial planning, emphasizing the importance of staying informed about available financial resources through consistent budget review. This statement reflects a practical aspect of managing finances and contributes to understanding an individual's engagement in financial planning activities.

If respondents actively consider the steps they need to take to stick to their budget (PROPPLAN_2)

The variable "PROPPLAN_2" indicates that deliberate efforts pertaining to budget adherence. Respondents are asked to specify their responses to the statement. This article emphasizes a proactive approach to financial planning, stressing the deliberate effort to stick to a budget by thoughtfully analysing and organizing the required actions. A favourable response contributes to a more comprehensive picture of an individual's involvement in prudent financial planning techniques by demonstrating a purposeful and intentional commitment to continuation financial discipline.

If respondents set financial goals for what they want to achieve with their money (PROPPLAN_3)

Financial goal-setting is the subject of the variable "I set financial goals for what I want to achieve with my money (PROPPLAN_3)". It is requested of respondents to indicate how much they agree or disagree with this statement. On a scale a response of 5 suggests a strong agreement to the statement and 1 indicates disagreement with the statement. Having specific financial goals and taking a thoughtful approach to managing finances are indications of a satisfactory reaction. By establishing assessable objectives, this statement advances our prediction of facing material hardship in life of an individual.

Measures of financial management (MANAGE1_1 to MANAGE1_4)

These variables encapsulate a range of essential financial behaviours, prompting respondents to express their agreement or disagreement with statements regarding their financial habits. This includes aspects such as timely bill payments, adherence to budgetary plans, settling credit card balances in full each month, and regularly inspecting financial documents for potential errors. Collectively, these behaviours provide a comprehensive view of an individual's financial responsibility, discipline, and diligence. A positive response indicates a positive and painstaking approach to managing finances, covering key features of budget adherence, credit management, and financial document scrutiny. Inclusion of these variables will provide a better approach to predict my target variable better.

Savings Habit of Respondents (SAVEHABIT)

The respondent's attitude toward saving money is evaluated by the "SAVEHABIT" variable, which asks them to grade the statement "Putting money into savings is a habit for me." Responses range from 1 to 6, with 6 indicating a robust agreement. This variable assesses the degree of a person's commitment to setting aside money on a regular basis, offering valuable information about their spending patterns and perspectives on creating a savings regimen. An insight from an individual’s savings tendency it can be predicted if an individual will suffer from material hardship or not.

Respondents sense of re-using an item (FRUGALITY)

The "FRUGALITY" variable, whose answers range from 1 to 6, assesses a person's attitude toward recycling goods rather than buying new ones. "If I can re-use an item I already have, there's no sense in buying something new" is a statement with which a score of six indicates high agreement. This variable is envisioned to forecast material hardship and gives information about how thrifty the respondent is or whether they prefer to reuse their current possessions over making new ones. A higher dedication to economical living is indicated by higher scores, which may lessen the chance of experiencing material hardship.

If respondents do own research before making decisions involving money (ASK1_1)

The purpose of the "ASK1_1" variable is to measure the respondent's research activity prior to financial decision-making. Contributors are asked to select an answer on a 5-point rating system, where 5 represents "Always" and 1 represents "Never." This measure is a valuable predictor of material difficulty since it shows how practical the person is in obtaining information and doing study prior to making financial decisions. Based on this scale, a higher score indicates a stronger tendency to regularly conduct research, which might lead to more cultured financial decisions and lower the risk of suffering material hardship.

If respondents ask other people their opinions before making decisions involving money (ASK1_2)

The variable "ASK1_2" is intended to assess the respondent's behaviour in seeking other people's views before making decisions concerning money. Respondents are asked to choose a response on a scale from 1 to 5, where 5 corresponds to "Always" and 1 to "Never." This variable is important in predicting material hardship as it offers understandings into the individual's inclination to seek external input and perceptions before making financial decisions. A higher score on this scale implies a greater tendency to consistently seek opinions, potentially reflecting a more collaborative decision-making approach and potentially swaying financial outcomes.

Individual’s proficiency in working with percentages? (SUBNUMERACY1)

The variable " SUBNUMERACY1" (How good are you at working with percentages?) is a variable used to predict material hardship. Participants are enquired to rate their proficiency on a scale from 1 to 6, where 6 corresponds to "Extremely good," and 1 indicates "Not good at all." This variable aims to capture individuals' self-assessed competence in handling percentages, and it contributes to predicting material hardship by providing perceptions into their numerical literacy and comfort with percentage-based calculations. Higher scores on this scale may propose greater confidence in working with percentages, potentially inducing financial decision-making, and extenuating the risk of material hardship.

If respondents belief that ability to manage money is NOT changeable (CHANGEABLE)

The "CHANGEABLE" variable collects respondents' opinions on how flexible someone's financial management skills are. Survey participants are invited to identify their level of agreement or disagreement using a seven-point grading system, where 7 signifies "strongly agree" and 1 represents "strongly disagree." The purpose of this variable is to measure opinions regarding the adaptability of financial management abilities. Higher scores on this scale, particularly in the "disagree" category, indicate a conviction that one's financial management skills are modifiable and improvable. This variable can be a key metric to predict material hardship of an individual.

Financial Knowledge (FINKNOWL_1 to FINKNOWL_3, KHKNOWL1 to KHKNOWL9.)

The purpose of the 'Financial Knowledge' variable is to assess respondents' comprehension of essential financial concepts, with a focus on predicting material hardship. It comprises of several questions on compound interest, the impact of inflation on savings returns, and the distinctions in volatility between stocks and mutual funds, aiming to gauge the extent of respondents' financial literacy. A higher score on this variable specifies a more profound understanding of basic economic and investing principles, potentially leading to better financial decision-making. The correctness of answers contributes to the score, with a perfect score earning a value of 3 and a null value assigned if the respondent is unable to answer any question correctly. Questions about long-term investment returns, differences in stock, bond, and savings volatility, diversification benefits, potential stock market losses, life insurance understanding, risks related to housing market investments, credit card payment nuances, bond-interest rate relationships, and the consequence of erm duration on the entire interest paid are all included in this assessment of financial literacy. More erudite understanding of economic and investing concepts is designated by a higher aggregate score on this variable. This might enable people to make better financial decisions and, as a result, lower their risk of experiencing material hardship. Therefore, a grasp of financial literacy provides insights into an individual's materialistic well-being. These variables have been referred in Lusardi and Mitchell's 2008 paper, "Planning and Financial Literacy: How Do Women Fare?" and detailed in Knoll and Houts' 2012 study, "The Financial Knowledge Scale: An Application of Item Response Theory to the Assessment of Financial Literacy."

If any household member received SNAP benefits (SNAP)

The "SNAP" variable can be utilized as a material hardship predictor and is formatted as a single-select question. Individuals were asked to respond if they or any family member had aid from the Food Stamp Program or SNAP (Supplemental Nutrition Assistance Program) in the earlier year. The positive response (1) advocates that a minimum of one member of the family was receiving SNAP benefits, which might be construed as a sign of financial worries or need among the respondents. This can be a key metric to predict material hardship of an individual.

Experience with Financial Agencies (CONSPROTECT1 to CONSPROTECT3)

The variable "Experience with Financial Agencies" incorporates critical aspects of peoples' engagements with financial services. To provide perception on the capacity of these exchanges, it also provides an evaluation of how frequently respondents felt disrespectful or abused in their financial transactions. The variable also inquiries about the respondents' acquaintance with agencies and organizations that are proposed to resolve financial services-related difficulties, indicating their awareness with the support systems that are available. It also looks at the respondents' level of engagement in pursuing answers by defining if they have actively reported issues to these agencies or groups. All together, these foundations provide a sophisticated perspective on people's experiences, obstacles, and proactive steps when interacting with the financial services industry**.** Thus, these queries will provide an insight to the prediction of material hardship of an individual.

If Respondents apply for credit or not (REJECTED_2)

The "REJECTED_2" variable indicates if respondents decided not to apply for credit because they thought they would be denied. With the purpose of revealing people's opinions about their solvency and possible fears of rejection, this variable seeks to comprehend why people make the choices they do about applying for credit. If the response is affirmative (1), it means that the respondent decided not to apply for credit because they were anxious or expected to be denied. Gaining an understanding of these dynamics aids one to better understand people's financial habits and credit-related decision-making processes. This variable can be influential in predicting material solvency of an individual.

Whether a victim of financial fraud or attempted financial fraud in past 5 years (FRAUD2)

To predict material hardship, the "FRAUD2" variable asks respondents if they have experienced financial fraud in the previous five years or if they have ever been the victim of financial fraud. Because it offers evidence on the frequency of financial fraud incidents—a factor that can have a substantial influence on an individual's financial stability—this variable is expressly valuable in forecasting material hardship. A response of (1) highlights the significance of comprehending the influence of fraud events on the general financial well-being of the questioned community by implying a possible correlation with material hardship.

If respondents lost a job (SHOCKS_1)

The "SHOCKS_1" variable focuses on whether respondents have lost their jobs and acts as a predictor for material hardship. This variable plays a key role in material hardship prediction as it accounts cases of job loss, a major life event that can have a significant effect on financial stability. A positive response (1) rises the possibility of a relationship with material hardship and serves as a crucial indicator of the financial problems that members of the surveyed group may be facing. Thorough assessments of material hardship must consider the effect of job loss on financial results.

Living Status at the age of 17? (HSLOC)

By inspecting respondents' living conditions when they were 17 years old, the "HSLOC" variable is used to predict material hardship. This variable is beneficial for material hardship prediction because it may divulge information about early life circumstances and geographic regions, which might affect different elements of financial well-being. Answers to this question further knowledge on the relationship between historical living circumstances and present or potential financial difficulties.

Highest level of education by person/people who raised respondent (PAREDUC)

The "PAREDUC" variable measures the highest level of education achieved by the individuals who raised the respondent, offering noteworthy insights into the respondent's background. This variable is of implication as it provides a demographic snapshot of the educational background of those who played a pivotal role in the respondent's early life. It obliges as a key factor for understanding potential influences on the respondent's educational and economic outcomes, with higher levels of parental education often correlating with positive impacts on various aspects of an individual's life trajectory. In this response 1 indicates lower level of education and 6 represents a graduate degree.

Diligence of Respondents toward long-term goals (SELFCONTROL_3)

The "SELFCONTROL_3" variable gauges how well respondents have faith in they can work hard to achieve long-term objectives. This variable is intended to measure people's self-control and persistence in working toward long-term goals. The score represents frequency from ‘Not at all’ to ‘Completely well.’ Higher scores on this scale echo a better sense of self-perceived capacity to work diligently toward goals, which may provide information about respondents' discipline and dedication to long-term goals.

If Respondent financially supports children (KIDS_NoChildren)

The "KIDS_NoChildren" variable is used to predict material hardship according to respondents' debt for their kids. This variable is vital to grasping children's financial responsibilities and how they may affect material hardship of an individual. An answer of "0" denotes that the responder does have children and provides for them financially, whereas a response of "1" denotes that there are no financial support obligations for children. Examining this variable helps to provide a more comprehensive assessment of the relationship that exists between monetary hardship and parental duties in the group under study.

Employment Status (EMPLOY)

Based on respondents' primary or only job status, the "EMPLOY" variable is proposed to predict material hardship. An examination of the relationship between various work statuses and the probability of facing material hardship is made possible by this variable, which offers a thorough summary of respondents' employment circumstances. A response of 1indicate the individual to be self-employed and highest response of 8 suggest the person has retired from work. Thus, a thorough response of this query might lead to a valuable insight regarding prediction of material hardship in an individual.

Age of Respondents (agecat)

The "agecat" variable is designed to predict material hardship based on respondents' age categories. This variable provides a segmented view of respondents' age groups, allowing for an analysis of the relationship between different age categories and the likelihood of experiencing material hardship. The specified age ranges provide a inclusive depiction of the respondents' age distribution, enabling researchers to discover potential patterns or trends related to material hardship across various stages of life. The age group ranges from 18 years to 75 years. The variable has been segregated in different categories in this range. An indication of different age groups vulnerability to material hardship will provide an idea of how material hardship is experienced by different individual.

Race/ Ethnicity of Respondents (PPETHM)

The "PPETHM" variable is projected to predict material hardship based on respondents' race/ethnicity. This variable facilitates the investigation of any possible relationship between the probability of facing material hardship and various racial and ethnic backgrounds. It can be investigated through any differences or trends in material hardship between different demographic groups by grouping respondents into discrete racial and ethnic categories. The categories listed here are, White, Non-Hispanic, Black, Non-Hispanic, Other, Non-Hispanic and Hispanic.

Gender of Respondents (PPGENDER)

The "PPGENDER" variable captures respondents' gender, categorizing them into two groups, Male, Female. The aforementioned variable provides a fundamental yet critical demographic feature that may be employed to inspect possible variations in understandings of material hardship between genders. A more nuanced knowledge of how financial issues may differ between male and female respondents within the studied population may be gained by examining material hardship via the gender lens. The recorded responses are 1 as Male, 0 as Female. It can lead to a valuable insight surrounding the topic of predicting material hardship of an individual.

Household Size of Respondents (PPHHSIZE)

The "PPHHSIZE" variable represents household size, with response values ranging from 1 to 5+. The number of people residing in the respondent's household is measured by this variable. Each number in the range 1 suggests a single-person home, while 5+ reflects a bigger household with five or more individuals. Because larger families may have dissimilar financial dynamics than smaller ones, analysing household size can provide insights into how the number of residents in a home may connect to or influence experiences of material hardship.

Household Income (PPINCIMP)

Household income is denoted by the "PPINCIMP" variable, which is divided into several income brackets. Exploring the correlation between household income levels and experiences of material hardship is made possible by this variable. Examining material hardship in relation to various income groups sheds light on how monetary complications could change depending on the home economic situation of the respondents. The income criteria start from less than $20,000 comprising of total 9 categories up to income of $150,000 and more.

Marital Status (PPMARIT)

The "PPMARIT" variable is used to categorize respondents into distinct marital status groups based on their marital status. A scrutiny of the relationship between experiences of material hardship and marital status is made possible by this variable. Analysing material hardship in connection to marital status sheds light on possible benefits or financial drawbacks of various relationship configurations. The responses include Married, Divorced/Separated, Widowed, never married, Living with partner

3.3. Data Preprocessing

3.3.1. Data Cleaning

High-quality data is essential to guarantee that the data set is apt for use in actual financial intervention decision-making. Inadequate or skewed predictions might be produced by machine learning models that are trained on subpar data. This goal is aided by data cleansing.

The original dataset I am using is already quite inclusive and has already been cleaned by the CFPB once. Hence, my data cleaning process will mostly comprise of the following steps: removal of missing values, removal of duplicates, data type conversion, and removal of rows with no response values.

3.3.1.1. Removal of Missing Values

Majority of machine learning models lack the intrinsic capacity to manage missing values. Tree-based models like random forests and decision trees are a couple of the exceptions. As a result, before inputting the data to the machine learning models, the imputation or elimination of missing values is a critical step in the data preparation process.

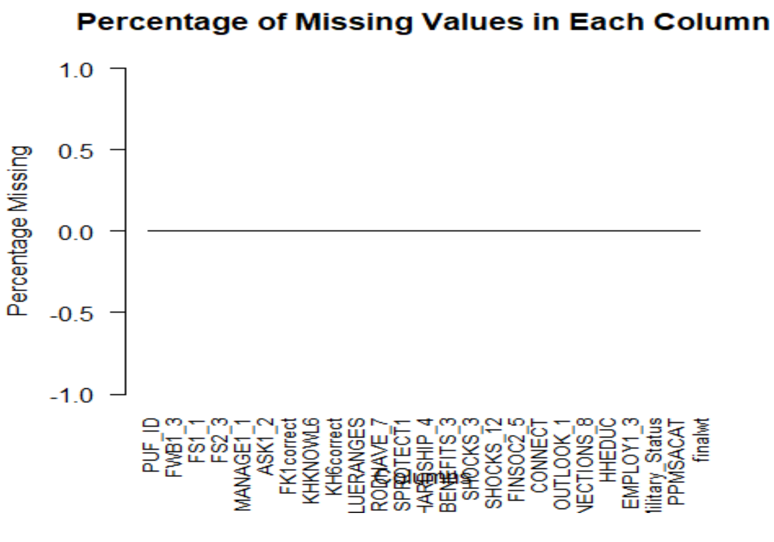


Figure: Percentage of Missing Values in Column

3.3.1.1. Modification of the Target Variable

My target variable consists of variables from MATHARDSHIP_1 to MATHARDSHIP_6. I am adding all the variables to make it into one single variable titled 'Hardship.' This process will help me to build my model around this specific target variable. I will tend to remove values which do not hold any significance and will not be adding any value in my model. So, participants who refused to answer to any of the questions asked I will discard them to achieve my desired target variable. Modifications of my desired variable reduced my dataset to 6365 rows and 217 columns. The analysis of the dependent variable showed that dataset has an imbalanced nature where majority of the participants on the lower side of the response. I will label my dependent variable in two categories; Facing Major Hardship, Not Facing Hardship. Individuals who have a response of 9 or greater than it on a scale of 18 are considered to be in the risk of facing a financial crisis.

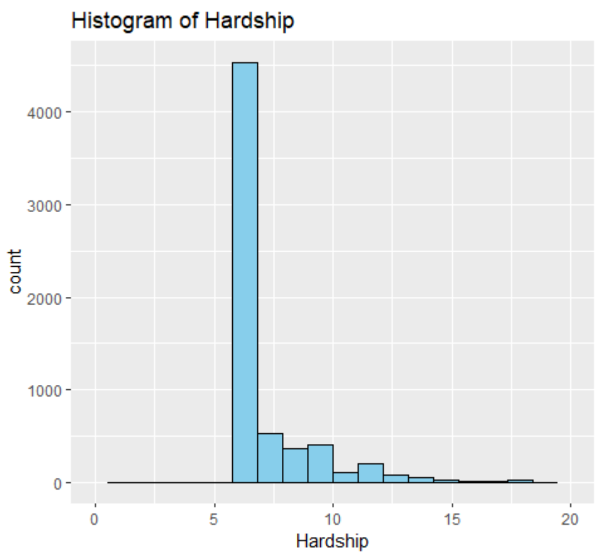


Figure: Histogram of Target Variable

3.3.1.1. Removal of Values with No Response

Resolving missing data is important to preserving the validity of my findings during the analytical stage of my research. To guarantee that the analysis was found on comprehensive and significant data, it was decided to exclude observations that lacked replies.

3.3.1.2. Removal of Unique ID Column

As a discrete primary key, the "PUF_ID" column has no predictive implication. As each instance has a unique value, the machine learning models are unable to recognize any trends. Between this column and the target column, there will not be any association at all.

3.3.1.3. Data Type Conversion

For most machine learning models to use data kinds of columns properly and efficiently, they must be converted to the suitable format. The models may interpret data in a biased way if the incorrect type of data is used. 5 prime categories may be used to classify most data: numerical, categorical, time-series, text, and picture data.

All survey item-related columns in my dataset will be transformed into categorical data types. For memory use, categorical data types outperform object or text data types. Because it minimizes the memory footprint, this is especially accommodating when working with large datasets. Compared to object data types, operations on categorical data can be performed more rapidly. In some kinds of analysis and procedures, this can result in better performance. It is common practice to prepare the dataset for supervised learning by converting relevant columns to the "category" data type. This supports in guaranteeing that the data is configured appropriately for a range of machine learning methods. (Exhibit 1)

3.3.1.4. Removal of Survey Items with Incomplete Base

Items in the survey with incomplete bases are eliminated. For each survey item, the term "Base" represents the person who was questioned. Therefore, in order to ensure that the questions apply to all respondents and not just a subset of those who provided pertinent answers to earlier questions, I only kept features where Base = "All".

3.4. Data Transformation

In the process of formulating the dataset for predicting material hardship, I presented a custom label encoder, 'MyLabelEncoder', which extends the abilities of the scikit-learn LabelEncoder class. This custom encoder precisely overrides the fit function to prevent the sorting of classes during the encoding process. This modification is substantial while dealing with categorical target variables, ensuring that the order of classes is preserved, which may hold significance in specific scenarios.

Following the customization of the label encoder, I omitted the target variable placed at index 72 from the feature set ('X'). This variable, representing material hardship, is essential for supervised learning. By excluding it from the features, I ensured that the model is not trained on the variable it aims to predict, averting data leakage.

For the remaining columns, I used one-hot encoding to further prepare the dataset for analysis. By means of this method, categorical variables are extended into binary columns, which provides a format that is fitting for a range of machine learning methods. Binary representations of the categorical features are added to the final dataset. One-hot encoding retains all the information present in the original categorical variable. Each category is represented by a unique binary column, preserving the uniqueness of each category.

I then used the custom label encoder to describe and encode the target variable. The target variable was encoded with numeric values ("0" for "Not Facing Hardship" and "1" for "Facing Major Hardship"), which characterize an individual's level of difficulty. This encoding offers the required structure for supervised learning tasks, which is critical for training a predictive model. The resulting binary columns make the encoding exceedingly interpretable.

These transformation steps lay the foundation for a predictive modelling task focused on material hardship. The customization of the label encoder maintains the integrity of categorical classes, and one-hot encoding confirms that the features are correctly formatted for machine learning analysis. This scrupulous preparation is involved in making a model capable of successfully predicting material hardship based on the selected features in the dataset.

3.4 Data Partitioning

To evaluate my machine learning models using a split train-test approach, data partitioning is needed. This reduces the likelihood of the overfitting problem. The models would have perceived all the data and would have produced unduly expectant performance scores if I had trained and evaluated them on the same set of data. To get an unswerving estimate of the classification accuracies of my models, an independent assessment of them on unseen data necessitates the use of a different set of data. It was necessary to split my data set into two subsets for the procedure: the training dataset and the testing dataset. For the split ratio, I used the most used 70:30 train-test split. This indicates that 70% of the data will be used for training the machine learning models while 30% of the data will be used to evaluate the performance of the models. (Exhibit 2)

4. Model Development and Evaluation

4.1. Model Development

4.1.1. Choice of Machine Learning Algorithms

The analytics strategy and issue being addressed determine which machine learning algorithms should be used. A strategy based on supervised learning would be suitable, as it is necessary to predict which candidates fall into the categories of "Not facing hardship" and "Facing major hardship," and the labels have already been attained by data processing. The fact that I am predicting categories rather than continuous values further identifies the prediction issue as a classification task. As a result, classification tasks should be supported by the machine learning algorithms that are selected.

Over the years, numerous machine learning algorithms have been developed, and each one has pros and cons. Predictability and interpretability are typically trade-offs. Models that are straightforward to compute, linear, and have well established connections are highly interpretable. Non-linear correlations and lengthy calculation times are typical features of high predictive power models.

Four distinct algorithms are selected in order to create a balance between the exploration of adequate machine learning methods and the avoidance of undue computing time. These four types of classifiers include logistic regression, Bernoulli naive Bayes, decision tree, and support vector.

Logistic Regression Model: The logistic regression model predicts the likelihood that an occurrence will fall into a specific category. A probability value between 0 and 1 is obtained by transforming the output of the linear combination of characteristics using the logistic function (sigmoid function).

$$P(y=1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \ldots\ldots + \beta_n x_n)}}$$

Here, P(y=1) is the probability of facing material hardship**.** β_0 is the intercept. β_1, β_2, β_3, β_n are the coefficients of independent variables x_1, x_2, x_3, x_n. In my model, the target variable is predicting material hardship, this equation captures the effects of independent variables on my desired target variable. (Exhibit: 11)

Non- Linear Support Vector Model: A supervised machine learning technique used for regression and classification problems is Support Vector Machine (SVM). SVM operates by determining the best hyperplane in the feature space to divide various classes. Support Vector Machines (SVMs) can capture both linear and non-linear relationships in data.

The SVM in my model is a non-linear classifier as the kernel I have set is "rbf" (radial basis function). The SVM will consider non-linear correlations between features because of the Radial Basis Function kernel, which aids the SVM to map the input data into a higher-dimensional space**.** Each support vector in SVM is given a weight, also known as a dual coefficient. The relevance of each decision-making point is represented by these coefficients.

Support Vectors: [[0. 0. 0. ... 0. 1. 0.]

[1. 0. 0. ... 0. 1. 0.]

[0. 0. 1. ... 0. 0. 1.]

...

[0. 0. 1. ... 0. 1. 0.]

[0. 0. 0. ... 0. 1. 0.]

[1. 0. 0. ... 0. 0. 1.]]

Dual Coefficients: [[-0.57705421 -0.57705421 -0.57705421 ... 0.60685098 0.60105096

1.00062346]]

Here, the support vectors are precise data points from the training set that play a decisive role in defining the decision boundary of the SVM. Each row in the matrix signifies a support vector. Each column represents to a feature my dataset. The dual coefficients are related with each support vector. They show the importance or contribution of each support vector to the structure of the decision boundary.

Bernoulli Naive Bayes Classifier Model: Naïve Bayes classifier is a supervised learning algorithm based on Bayes theorem. This is generally used for solving classification problems. The primary use of this model is in text classification that includes a high-dimensional training dataset. Assuming independence between features according to the class, naive Bayes classifiers, such as Bernoulli Naive Bayes, are fundamentally linear models. While the "naive" assumption of independence makes modelling easier, it also makes it difficult to visualize intricate, non-linear correlations in the data. The model operates on Bayes theorem.

$$P(Y \mid X) = \frac{\mathrm{P}(\mathrm{X} \mid \mathrm{Y}).P(Y)}{P(X)}$$

Here, P (Y|X) is the probability of facing material hardship given the values of the independent variables. P(Y=1) is the prior probability of facing material hardship. P (X|Y) is the likelihood of observing the value of 'X' given that material hardship is true.

For Bernoulli Naïve Classifier, the operation follows a certain pattern.

$$P(Y = 1 \mid X_1, X_2, X_3, \dots) \propto P(Y = 1) \cdot \prod_{i=1}^{n} P(X_i = x_i \mid Y = 1)$$

Here, $P(Y{=}1)$ is the prior probability of facing material hardship, and $P(Xi{=}xi|Y{=}1)$ characterizes the likelihood of observing the value xi for variable Xi given that dependent variable is true. The probability $P(Xi{=}xi|Y{=}1)$ is naturally modelled using a parameter θi limiting from i_1 to x_i, which embodies the probability of observing $Xi{=}1$ (presence of variable) given that material hardship is true.

$$P(Xi = xi \mid Y = 1) = \theta \cdot (1 - \theta i_1)^{(1-xi)}$$

For Bernoulli Naive Bayes, the model estimates these parameters θi_1 for each variable Xi based on the training data. The probability for each variable is then combined using the product formula in the expression above.

Decision Tree Model: Decision Trees are adaptable and natural-feeling machine learning models that simulate human decision-making. It can decompose complicated topics into a sequence of easier options. Decision Trees intrinsically capture non-linear relationships in the data. Unlike linear models, they can create complex decision boundaries by combining simple rules.

For the hyperparameters I have used in my model:

- **Criterion:** 'entropy'
- **Max Depth:** 3

- **Min Samples Split:** 40
- **Min Samples Leaf:** 2
- **Max Features:** 'sqrt'
- **Min Impurity Decrease:** 0.0005
- **Class Weight:** 'balanced

The decision tree is constructed with a maximum depth of 3, meaning it can make decisions based on up to three variables. The 'entropy' criterion is used to measure the quality of a split. The tree will not split further if the number of samples at a node is below 40. The top node is signified as root node, the tree moves to left child node if the decision is true and moves to right node if the decision is false.

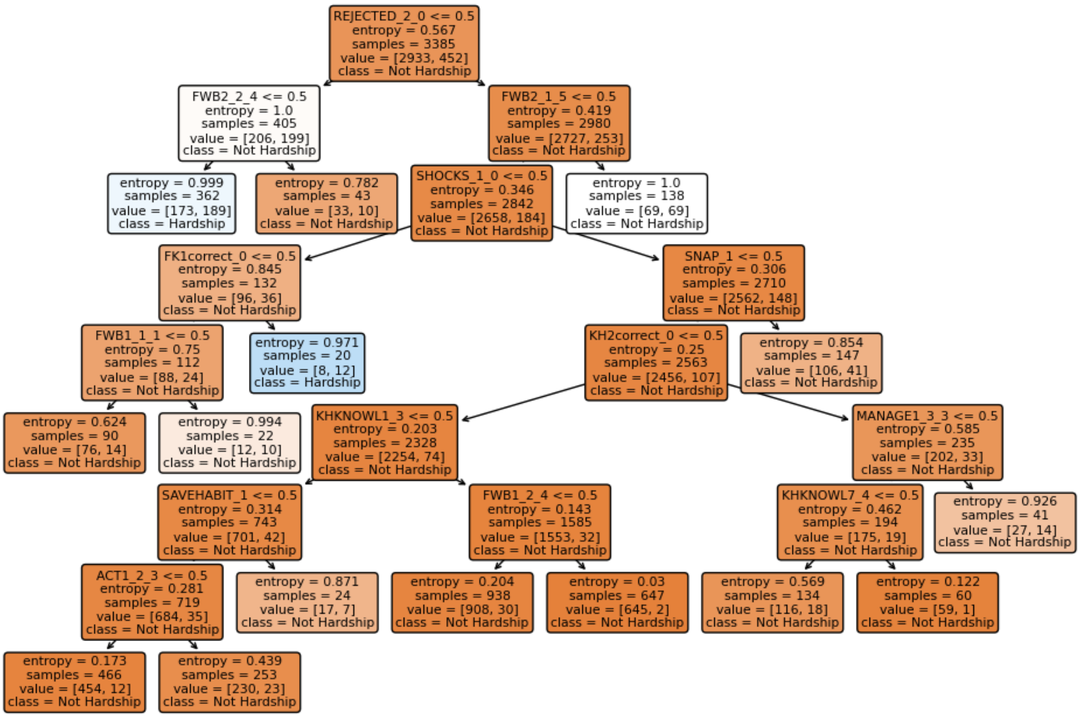


Figure: Decision Tree Model's Plot Tree

4.1.2. Model Evaluation

While assessing the trained machine learning models, selecting the assessment metrics is a crucial consideration. I will look at four important assessment measures that practitioners frequently employ. These are F1-Score, recall, accuracy, and precision.

Since accuracy is easy to compute and comprehend, it is perhaps the most often used evaluation statistic. It is basically the percentage of predictions that the machine learning model accurately made. Its boundaries fall between 0 and 1, where 0 represents the lowest score and 1 represents the highest. When the class distribution is balanced and the true positives and true negatives are more significant, the accuracy score is typically employed.

The precision measures how many genuine positives there are among all the anticipated positives. When the expenses of false positives are extensive, this assessment metric is suitable.

The percentage of true positives among the total number of genuine positives is called recall. When the costs of false negatives are substantial, this assessment metric is appropriate.

The harmonic mean of recall and accuracy is known as the F1-Score. When the expenses associated with false positives and false negatives are both significant, this assessment measure is fitting. The F1-Score is also superior to the accuracy score for unbalanced data sets. In its computation, it disregards real negatives. Because the machine learning model can always predict the majority class, a data set with a large proportion of one class typically has an inflated accuracy score. The accuracy of the baseline model, which always predicts 1, is equal to the marginal probability P (y = 1), or 0.14. This is because in my dataset 14% (682 responses out of 4919) of the labels are in the positive class and will be accurately identified. Since 86% (4237 responses out of 4919) of the labels are in the negative class, they will inevitably be incorrectly categorized. To be more precise, there will be 0.14 real positives and 0.86 false positives.

Baseline F1-score = 2 * (precision * recall) / (precision + recall)

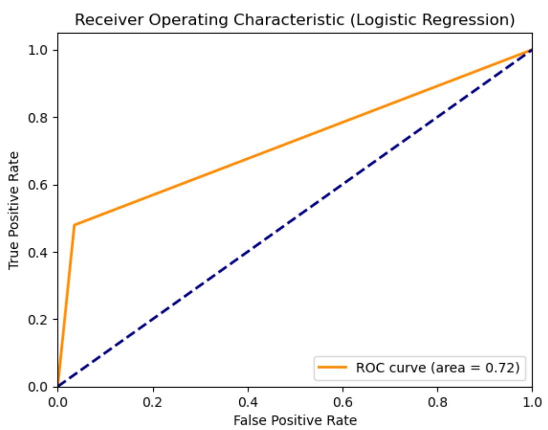
Since precision = 0.14 and recall = 1,

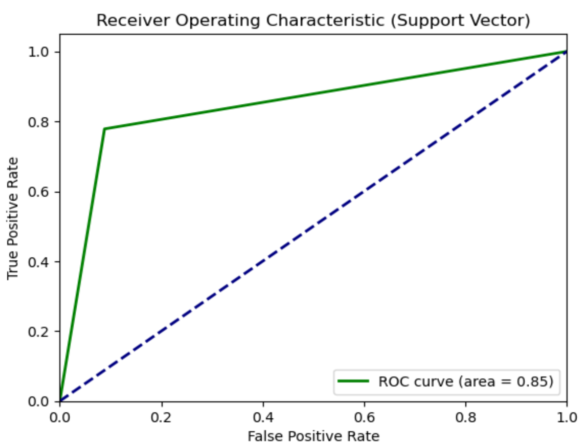
Baseline F1-score = 2 * (0.14 * 1) / (0.14 + 1)

Baseline F1-score = 0.28/ 1.14 = 0.25

Therefore, the benchmark F1-Score is 0.25 and my preferred model is required to outscore this score.

The efficiency of the model is also shown by the ROC AUC score. The model's ability to discriminate between the positive and negative classes is improved with a higher AUC.





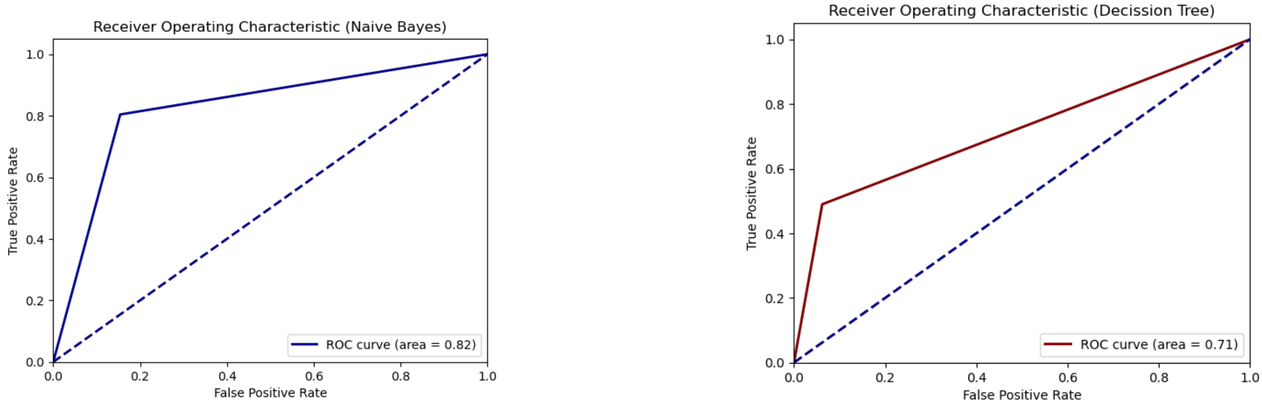


Figure: ROC- AUC curve of Models

As a result of my data set's imbalance nature, I will select the final machine learning model using the F1-Score assessment measure and higher ROC-AUC score. All four-assessment metrics, however, will be computed to look at the trade-offs between various machine learning models.

Table: Comparison of Evaluation Metrics of all models

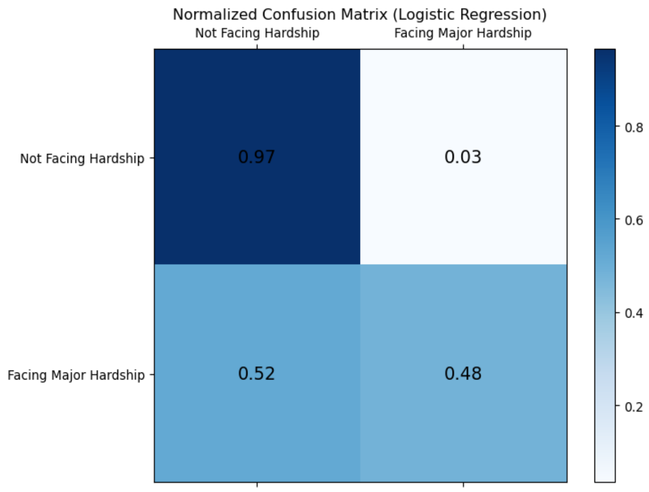
	Accuracy	**Precision**	**Recall**	**F1-Score**	**ROC-AUC Score**
Logistic Regression	0.9001	0.6788	0.4794	0.5619	0.72
Decision Tree Classifier	0.8595	0.4811	0.6546	0.5546	0.77
Non-Linear Support Vector Classifier	0.8939	0.5763	0.7784	0.6623	0.85
Bernoulli Naive Bayes Classifier	0.84802	0.4457	0.8041	0.5735	0.82

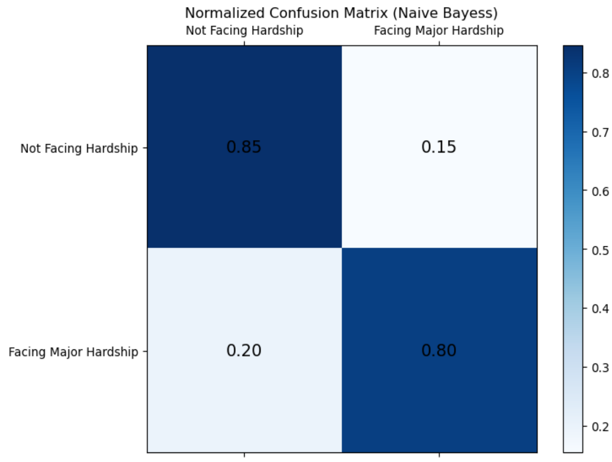
The accuracy, precision, recall, ROC-AUC Score and F1-Score for each of the four trained machine learning models are listed in the assessment metrics table above. Based on the test data set, all the scores were assessed. At 0. 9001, logistics regression has the greatest accuracy. With 0.8041, the Bernoulli Naive Bayes Classifier has the maximum recall value. At 0.6788, the logistics regression has the highest precision. With acing F1-Score of 0.6623 and ROC-AUC score of 0.85, the support vector classifier performs standard in all categories of the evaluation metrics.

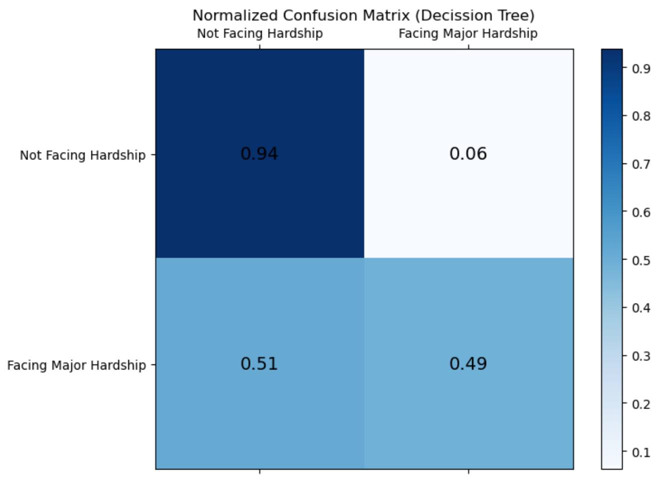
Also, I have analysed confusion matrix of all the models. A confusion matrix purposes similarly to a contingency table in terms of assessing how fine a machine learning model performs in classification. It is just a ‘N x N’ matrix or table, where N is the number of classes we are attempting to forecast. " Facing major hardship " and "Not Facing hardship," denoted by 1 and 0, respectively, are the two groups for our data set. Consequently, the table will contain two rows and two columns with a total of four elements since the value of N is 2.

True positives and true negatives are properly classified predictions. A normalized scale containing only one true positive and one true negative suggests that the machine learning model is sound and has produced no prediction mistakes at all.

It indicates that all the " Facing major hardship " individuals for this project have been effectively identified, allowing for the planning of further financial assistance for them. No " Not Facing hardship " applicants have been mistakenly assigned to the " Facing major hardship " category.







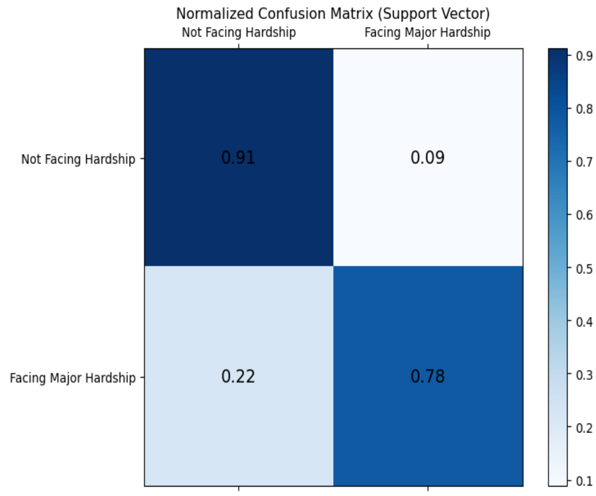


Figure: Confusion Matrix of all models

5. Results

5.1. Feature Importance in Logistic Regression

A metric acknowledged as "feature importance" divulges which variables have the major influence on a predictive model's predictions by indicating how much each feature contributes to the model. Feature importance is vital because it will help us to understand the factors that drive the model's performance by stressing on important variables that have a weighty effect on the outcome.

After I performed feature importance on the logistic regression model, I sorted out top 20 variables that are majorly influence the model's predictive capacity. The Financial Well Being query with the question 'Giving a gift...would put a strain on my finances for the month (FWB2_1)' has the most significant influence on the target variable. The variable with category (FWB2_1_5) of people who consider that offering a present would put strain on their finances are at the risk of facing major financial hardship. The variable comes along a coefficient of 0.6799 which indicates an increase in the value of variable FWB2_1_5 is linked with an increase in the likelihood of material hardship. Also, experience with financial agencies seem to have an impact on the prediction of the target variable. But it influences the result in opposite direction of what financial well-being does. A scenario where an individual has never experienced being mistreated decreases the individual of being a sufferer of material hardship. The CONSPROTECT1 variable is valued at -0.6100 which suggests that an increase in this criterion will lead to a decrease in the likability of the target variable being '1'. On the other hand, the respondents who do not feel the same, or have an experience of being mistreated by financial agencies falls into the category CONSPROTECT1_4, and the variable comes along with a coefficient of 0.609029 indicating that an increase or people falling into this criterion are in the risk of facing material hardship in life.

The variable PPINCIMP which is actually a household's income size shows that it the category where the household has income of $100,000 to $149,999, having a high income is less likely to encounter material hardship. The coefficient of this variable is -0.4960 suggesting a increase in this criterion will decrease the possibility of an individual's chance of facing material hardship. The tendency of paying bills on time is also turning out to be an important feature in predicting the likelihood of facing material hardship. A coefficient of -0.4272 suggests that if an individual pays bills on time are not at the risk of facing hardship in life. This is a key metric as an individual's tendency of paying bills timely can track the outcome of a target variable. A shock experience of losing a job shows that it has a positive coefficient of 0.40930 which indicates that if an individual has faced the shock of losing a job may face material hardship in life.

Overall, the feature importance will give an idea of how the outcome variable is being influenced by the independent variables.

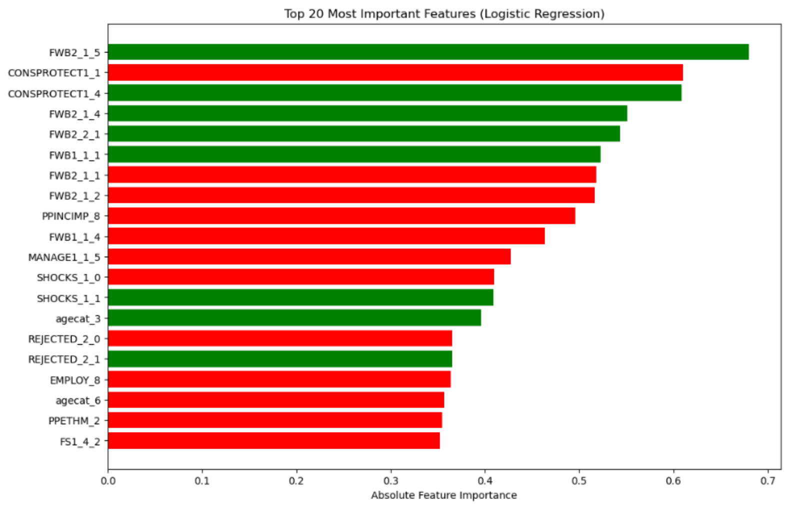


Figure: Feature Importance of Logistic Regression (Positive in Green, Negative in Red)

5.1.2. Odds Ratio in Logistic Regression

A statistical measurement called the odds ratio can be employed in logistic regression to evaluate the correlation between the probability of an event occurring and the existence of a given attribute. The likelihood of material struggle is positively impacted by features with larger positive odds ratios. There is a negative correlation between features with lower odds ratios and a decreased possibility of experiencing material hardship. By identifying the variables that considerably influence the frequency or avoidance of material hardship, understanding and utilizing odds ratios in my model analysis may help inform policy design, interventions, and decision-making. An odds ratio greater than 1 indicates a positive association. This means that the presence of the feature upsurges the odds of the dependent variable being 1.

The findings of odds ratio in my logistic regression model are quite like the findings of feature importance in the model. The individuals in the age category of "35-44" are in the risk of facing material hardship as the variable has a positive odds ratio of 1.49. A one-unit increase in this category is associated with 1.49 times higher odds of experiencing material hardship. On the other hand, individuals in age category of "62-69" have 0.70 times lower odds of experiencing material hardship compared to the reference category.

From odds ratio it can be inferred that a section of population who have been employed, but now is retired are not in the risk of facing material hardship. This section has a negative odds ratio of 0.69 which suggests that individuals in this criterion have less odds of falling into the category of facing material hardship in my logistic regression model. Similar to the feature importance of logistic regression the odds ratio also indicates that if a person has not faced any financial shoch i.e., losing a job is more likely not to face the crisis in life compared to an individual who has experienced such shock. The SHOCKS_1 variable with the criterion of losing a job has an odds ratio of positive 1.51 and the variable with category of not losing a job has an odds ratio of negative 0.66 which gives an insight that the person who has experienced such crisis is more in the risk of falling into the criteria of facing material hardship according to my logistic regression model.

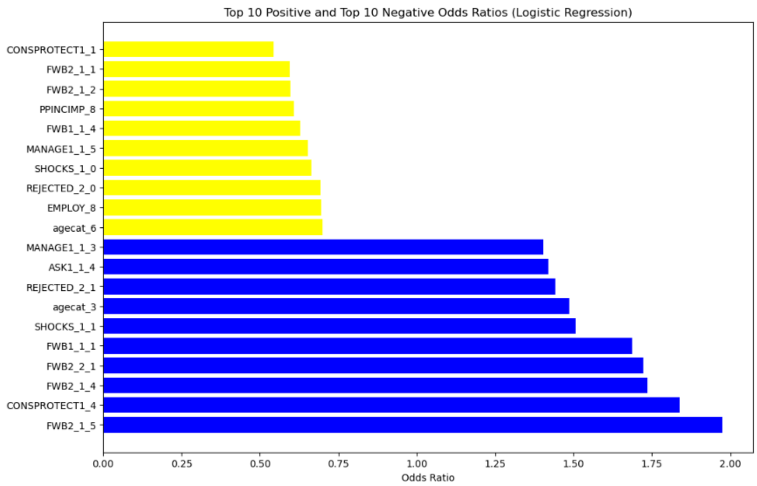


Figure: Odds Ratio of Logistic Regression (Positive in Blue, Negative in Yellow)

5.2. Feature Importance in Support Vector Model

In order to improve interpretability of the model, facilitate well-informed decision-making, and capitalize on model performance, feature importance analysis in predictive modelling is decisive. It helps in the identification of pertinent characteristics, gives a foundation for prioritizing actions, and offers perceptions into the major factors driving forecasts. Importance scores indicate the degree to which these features impact the SVM model's predictions. The features have a sturdy influence on the model's predictions. But it does not necessarily mean that feature alone leads to a prediction of 1. Changes in these features are likely to result in important changes in the model's output, making them vital for understanding the model's decision-making process.

Like logistic regression model the SVM model also tends to show that the variable FWB2_1 has the most influence on model’s prediction. This variable of category 5 has a feature importance value of 31.40 that suggests that responses to this inquiry have the maximum impact on predicting material hardship. A higher value for this feature meaningfully increases the model's prediction. Individuals who have responses to the question “I could handle a major unexpected expense” and respondents who have said they cannot at all handle a major unexpected expense are more in the risk of facing material hardship. This variable has a value of 26.90 that shows how influential this variable is to my model’s prediction.

Another financial well-being question “I have money left over at the end of the month” is also seen to have influence on the model’s predictive quality. Individuals with response of that they never have money left seem to be a critical influential variable in model’s decision-making progress as the feature importance value it has is 24.0. The management skills of finances of an individual also seem to be playing critical role in a model’s decision-making process. The MANAGE_1 variable which refers to the question that if an individual has paid his bills on time or not seem to have a great feature important value in SVM model’s prediction process.

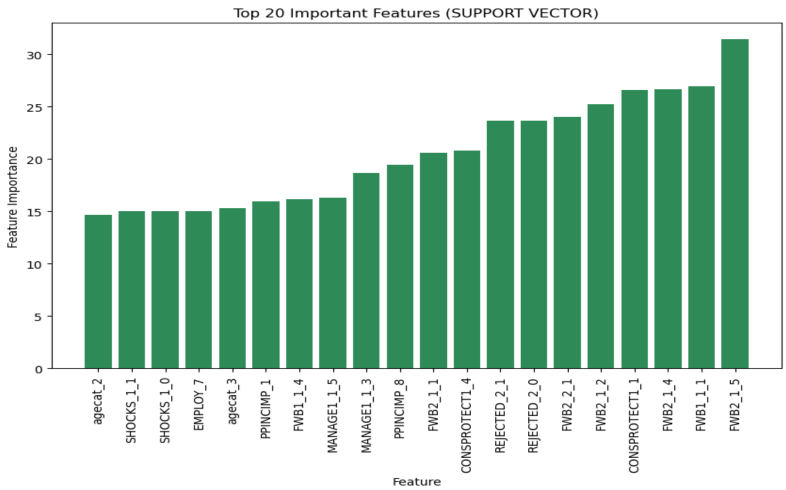


Figure: Feature Importance of Support Vector Model

5.2.1. Permutation Importance in Support Vector Model

Permutation importance evaluates the impact of each feature on model performance by randomly shuffling its values and measuring the resulting alteration in a specified scoring metric. Permutation importance quantifies the change in the F1 macro score (custom_f1_macro_score) when the values of each feature are randomly permuted. Higher values indicate greater effect on the model's performance.

Responses to MANAGE1_1 variable play a positive role in predicting material hardship, indicating that efficacious financial management practices contribute to the model's performance. Subsequently, the individuals who have been attentive in paying bills is a key metric to identify individuals who are in the risk of facing material hardship.

Interestingly, Both SNAP_0 & SNAP_1 have an importance value of 0.0074 indicating that both conditions—receiving and not receiving SNAP benefits—are associated with an increased likelihood of experiencing material hardship according to the model. Positive importance values show that when the values of these features are randomly shuffled, it leads to a decrease in the model's performance. In this framework, a positive impact suggests that the original, unshuffled values of "SNAP_0" and "SNAP_1" are telling for predicting material hardship.

The negative importance value of 0.0036 proposes that higher education levels (Some college/Associate) represented by "HHEDUC_3" are associated with a decreased likelihood of experiencing material hardship according to the model. The negative importance value suggests that belonging to ethnicity category 2, as embodied by "PPETHM_2," is associated with a decreased likelihood of experiencing material hardship according to the model. This category 2 indicates ethnicity of Black, Non-Hispanic. Negative importance values typically show a protective or extenuating effect on the predicted outcome. Specifically, individuals with either "FINGOALS_1" or "FINGOALS_0" are less likely to encounter material hardship compared to individuals without these goals or aspirations, according to the model's assessment.

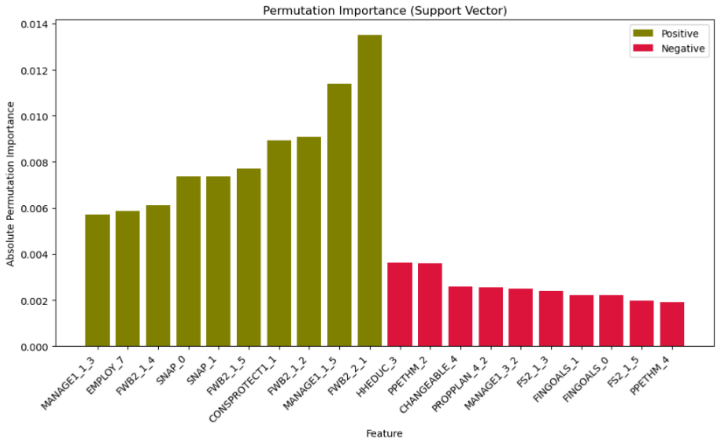


Figure: Permutation Importance of Support Vector Model

5.3. Log Probability Difference in Naive Bayes Model

Log probability difference offers valued insights into the variables driving the Naive Bayes model's predictions for material hardship, aiding in the understanding of main factors influencing the classification outcomes. A positive log probability difference for a feature suggests that the presence of that feature is more indicative of one class over the other, while a negative difference suggests the opposite.

Responses to variable SELFCONTROL_3, related to self-control behaviours, have a log probability difference of 1.61, making them a critical factor in predicting material hardship. The variable inquires this question- “I am good at resisting temptation.” Respondents who have responses that they do not have a good self-control can critically impact the model’s predicting behaviour. A person’s financial skill also play a vital role, indicated by variable

FS1_2 which corresponds to the query- “I know where to find the advice I need to make decisions involving money.” An individual responding to this query “Not at all” seems to be a key metric in the model.

Another financial skill query FS1_7 which inquires if a person knows how to make myself save also guides the model’s decision-making process. This variable has log probability difference of 1.86 which indicates a high log probability difference. The query of knowing how to keep myself from spending too much also contribute to the model’s prediction of material hardship in an individual. Two variables ACT_1 & ACT_2 relating to acting upon financial commitment are also substantial in the process of prediction through naïve bayes model. This shows both these variables have strong impact in predicting material hardship of an individual.

Responses to the frequency of paying bills timely play a critical role in the model’s decision-making process. The log probability difference of the MANAGE_1 variable is significantly high suggesting that this variable is noteworthy in the model’s prediction of material hardship in an individual.

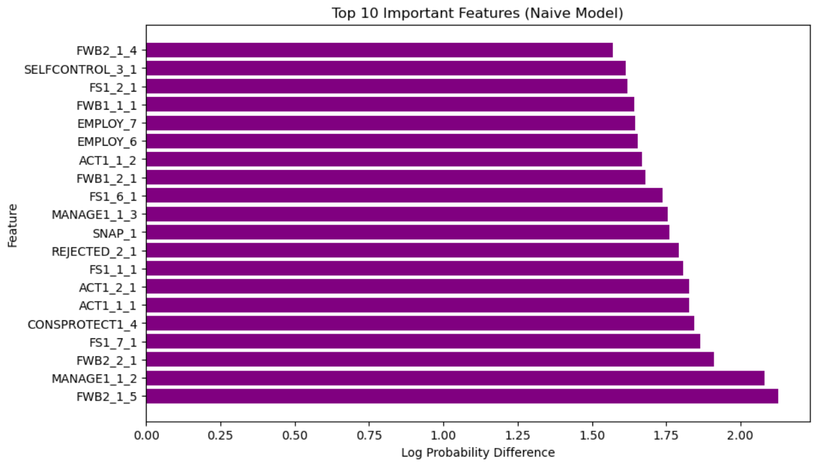


Figure: Log Probability Difference of Naïve Bayes Model

5.3.1. Permutation Importance in Naive Bayes Model

The effect of arbitrarily shuffling a feature's values on the model's performance is measured by permutation importance. Higher permutation importance values indicate that, when perturbed, have a more significant impact on the model's predictive performance. These features play a decisive part in the Naive Bayes model's capability to predict material hardship.

SNAP_1 and SNAP_0 characterize whether an individual has received SNAP benefits or not. FWB2_1 is a financial well-being question which asks if giving a gift would put a stress on the finances for the month. The equal permutation importance suggests that shuffling values for these variables has a consistent impact on the model, indicating their significance in predicting material hardship. Both SNAP responses have a coefficient value of 0.0026 and FWB2_1 also have a value of 0.0026 making these three variables most important in permutation importance criterion of naïve bayes model. Features like agecat_6, HHEDUC_3, PPHHSIZE, SUBNUMERACY1, MANAGE1_4, CONSPROTECT3, KH6correctnhave lower permutation importance, signifying a relatively lesser impact on the model's predictions.

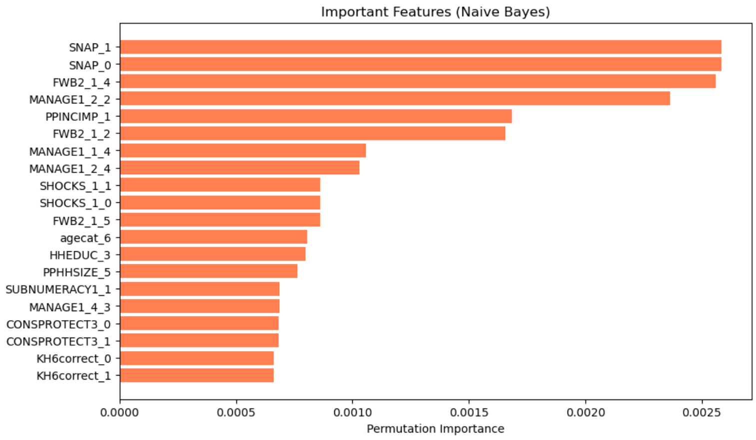


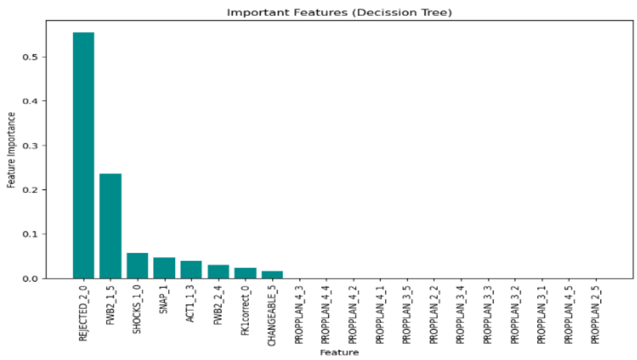
Figure: Permutation Importance of Naïve Bayes Model

5.4. Feature Importance in Decision Tree Model

Feature importance in a Decision Tree is a measure of how much each variable contributes to the model's decision-making process. Higher feature importance values indicate variables that are more critical in determining the outcome. A higher feature importance suggests that the presence or values of that feature are more indicative of the target variable being in category 1 rather than category 0.

REJECTED_2_0 is strongly associated with predicting the target variable to be in category 1. It has a feature importance value of 0.5541. An individual's perception on applying for credit is crucial in predicting the target variable. The variable of financial knowledge-based query if a person understands housing market losses is critical and indicate that an understanding of this issue proves to be vital in decision tree model's decision-making process. This variable comes along with a value of 0.1278.

Figure: Feature Importance in Decision Tree Model



Features like PROPPLAN_4, PROPPLAN_4, and others have importance values of 0.0000, indicating that these features do not contribute to decision tree model's decision-making process. They might not deliver much information in distinguishing between the two categories in the model's prediction process.

6. Discussion & Limitations

Material hardship is characterized by the incapability to fulfil basic needs, stands as a serious issue affecting individuals and families across diverse socio-economic backgrounds. The logistic regression model's feature importance analysis discloses that enquiries about financial well-being are crucial, with gift-giving-related factors having the most impact. The results of odds ratios indicate that household income and financial agency maltreatment are significant determinants. The concept stresses how perceptions of one's financial situation and the probability of suffering material hardship are intricately related.

The implication of financial well-being inquiries is highlighted by the SVM model, especially those that deal with unforeseen expenses. The model also stresses the importance of having money left over at the end of the month, paying bills on time, and having sound financial management abilities. This is consistent with the results of logistic regression, supporting the notion that proactive financial management is crucial in anticipating material hardship.

Due to my dataset's unbalanced nature using of kernel: rbf the support vector model can incorporate more complexities compared to other models.

Table: Model-wise evaluation of errors

	Logistic Regression	**Non-Linear Support Vector Classifier**	**Bernoulli Naive Bayes Classifier**	**Decision Tree Classifier**
True Positive	97 %	91%	85%	94%
True Negative	48%	78%	80%	49%
False Positive	3%	9%	15%	6%
False Negative	52%	22%	20%	51%

Also, as SVM can capture non-linear relationship, due to the nature of this dataset the SVM model is more suitable for further analysis.

The SVM's performance is comparatively balanced, with rates of 91% True Positive and 78% True Negative. False Positive rates for the SVM are 9%, while False Negative rates are 22%. Compared to other models, these rates are very balanced, yet there is still opportunity for improvement. The performance of SVM seems consistent across True Positive, True Negative, False Positive, and False Negative rates, signifying that it doesn't heavily favor one class at the expense of the other.

Feature Importance (Logistic Regression)	Odds Ratio (Logistic Regression)	Feature Importance (Support Vector)	Permutation Importance (Support Vector)	Log Probability Difference (Naïve Bayes)	Permutation Importance (Naïve Bayes)	Feature Importance (Decision Tree)
FWB2_1_5	FWB2_1_5	FWB2_1_5	MANAGE1_1_5	FWB2_1_5	SNAP_1	FWB1_1_1
FWB1_1_1	FWB1_1_1	FWB1_1_1	FWB2_1_5	MANAGE1_1_2	SNAP_0	MANAGE1_1_5
FWB2_2_1	FWB2_2_1	FWB2_1_2	FWB2_2_1	FWB2_2_1	FWB2_1_5	REJECTED_2_0
FWB2_1_4	FWB2_1_4	MANAGE1_1_3	SHOCKS_1_1	SNAP_1	KHKNOWL6_1	MANAGE1_1_4
EMPLOY_7	EMPLOY_7	FWB2_1_4	SHOCKS_1_0	REJECTED_2_1	KH6correct_1	FWB1_1_2
MANAGE1_1_3	MANAGE1_1_3	FWB2_2_1	FWB2_1_4	MANAGE1_1_3	KHKNOWL6_2	SNAP_1
REJECTED_2_1	REJECTED_2_1	REJECTED_2_1	FWB2_1_2	FS1_7_1	KH6correct_0	SHOCKS_1_1
SHOCKS_1_1	SHOCKS_1_1	REJECTED_2_0	agecat_2	ACT1_1_2	FWB2_2_1	KHKNOWL6_1
FS1_6_3	FS1_6_3	MANAGE1_1_5	EMPLOY_8	FS1_1_1	FS2_3_5	FWB2_1_5
agecat_3	agecat_3	FWB2_1_1	FWB2_1_1	FWB1_2_1	PPHHSIZE_5	KHKNOWL3_1
FS1_2_1	FS2_3_1	FWB1_1_4	FS1_6_3	EMPLOY_7	KHKNOWL7_1	FWB1_1_3
EMPLOY_8	PPINCIMP_8	EMPLOY_8	PPINCIMP_8	ACT1_1_1	FS1_3_1	FWB2_1_2
REJECTED_2_0	FWB2_2_5	EMPLOY_7	FWB1_1_1	FRUGALITY_1	FS1_2_1	PROPPLAN_4_2
SHOCKS_1_0	SHOCKS_1_0	PPINCIMP_1	MANAGE1_1_3	ACT1_2_1	KHKNOWL2_3	KH2correct_0
FWB2_2_5	REJECTED_2_0	PPINCIMP_8	REJECTED_2_0	FWB1_1_1	KH2correct_1	MANAGE1_4_5
PPINCIMP_8	EMPLOY_8	FS1_6_3	REJECTED_2_1	SHOCKS_1_1	KH2correct_0	FWB2_2_4
MANAGE1_1_5	MANAGE1_1_5	agecat_2	FWB1_1_4	SAVEHABIT_1	KHKNOWL2_1	PPGENDER_1
FWB1_1_4	FWB1_1_4	SHOCKS_1_1	EMPLOY_7	FS1_6_1	FRAUD2_8	EMPLOY_8
FWB2_1_2	FWB2_1_2	SHOCKS_1_0	PPINCIMP_1	MANAGE1_1_1	FRUGALITY_3	FRAUD2_0
FWB2_1_1	FWB2_1_1	MANAGE1_3_3	MANAGE1_3_3	EMPLOY_6	PPINCIMP_3	KH1correct_0

Table: Important Variables in different models (Red highlighted are common among models)

.

Although the SVM model is proven to be best among all models, after analysing all the model and extracting the variables that influence material hardship seem to be quite common among all the models. Variables concerning gift giving, saving tendency, managing finances, shock encounters are some of the common variables. The variables concerning income criteria of household in between $100,000 to $149,999 seem to be safe from experiencing material hardship. Unemployed or temporarily laid off individuals in criteria EMPLOY_7 and EMPLOY_8 prove to be influential in prediction process as individuals in this criterion are less in the risk of facing material hardship.

The Naive Bayes model offers sophisticated perceptions into the behavioural and psychological factors driving material hardship because of its importance on self-control behaviours and financial literacy. The focus on proactive money management techniques, such knowing where to get guidance and how to avoid temptation, is remarkable. The inspection of permutation significance offers insight into the recurring influence of specific demographic characteristics and SNAP answers.

The decision tree model highlights the critical role of perceptions regarding credit applications and understanding housing market losses in predicting material hardship. While certain variables, like PROPPLAN categories, exhibit lower importance, the model emphasizes the role of specific financial knowledge-based queries in the decision-making process.

Robustness of these predictors is reinforced by the uniformity of models in emphasizing the role of financial perspectives, management practices, and life events in predicting material hardship. Yet, the subtle variations in feature significance highlight how intricate the material hardship issue is, underscoring the necessity of a comprehensive strategy for forecasting. These models offer valuable insights that have significant policy ramifications. The most vulnerable people can benefit from interventions that encourage appropriate bill-paying habits, foster financial literacy, and deal with the effects of life events like losing one's job. Implementing systems to identify individuals who have recently experienced financial shock, such as job loss.

Providing targeted support services and resources to help the individuals navigate through challenging times. Stressing how vital it is to pay bills on time. Thinking of programs designed to inspire or reward people for continuing to pay their bills on time. Determining which age groups are most susceptible to financial stress and adjust solutions accordingly. Programs tailored to the specific age range of "35-44" persons, for example, may prove advantageous. Understanding how vital psychological elements like self-control and financial resiliency are. Taking action to promote mental health and wellbeing by developing improved coping strategies. Raising awareness about SNAP benefits and ensuring that eligible individuals are informed about and have access to these assistance programs. Promoting cooperation between governmental, nonprofit, and community groups to establish an all-encompassing support network involving private sector organizations in programs that improve financial security and lessen material hardship. Regularly updating predictive models by incorporating new data and evolving socio-economic indicators. This guarantees the accuracy and relevance of intervention strategies over time.

Policymakers, social workers, and community leaders need to collaborate to develop focused interventions that address the multifaceted factors discovered by the prediction models by combining various tactics. Interventions should be adaptive to the shifting dynamics of material hardship by being context-specific, adaptable, and guided by incessant assessment.

6.1. Limitations

Firstly, the predictive models are relying on specific datasets, and their generalizability to wider populations may be limited. If the dataset is not representative, the models may not capture the diversity of involvements related to material hardship.

Secondly, divergent perspectives among participants may exist over the precise meaning of different questions. Thus, this could have an impact on how "correct" their grading system is. Additionally, because this is a self-administered survey, the reliability of the results they offer may be somewhat limited.

Furthermore, the models classify relations between variables and material hardship but may not establish causation. Other unobserved factors could contribute to both the identified predictors and material hardship outcomes.

Although the models here reliable, there are some limitations to predict an individual's risk of facing material hardship. When compared to physical systems like machinery and equipment, individuals typically behave in an unexpected or random manner. Therefore, the machine learning model's future prediction may be partially invalidated as a result.

Potential problems include underfitting or overfitting. Model complexity must be balanced to guarantee consistent performance on fresh, untested data.

6.2 Policy Implementation & Future Work

While navigating the complex landscape of material hardship, policy implementation recommendations and future work considerations is crucial to contribute to ongoing efforts in building a more resilient and equitable society.

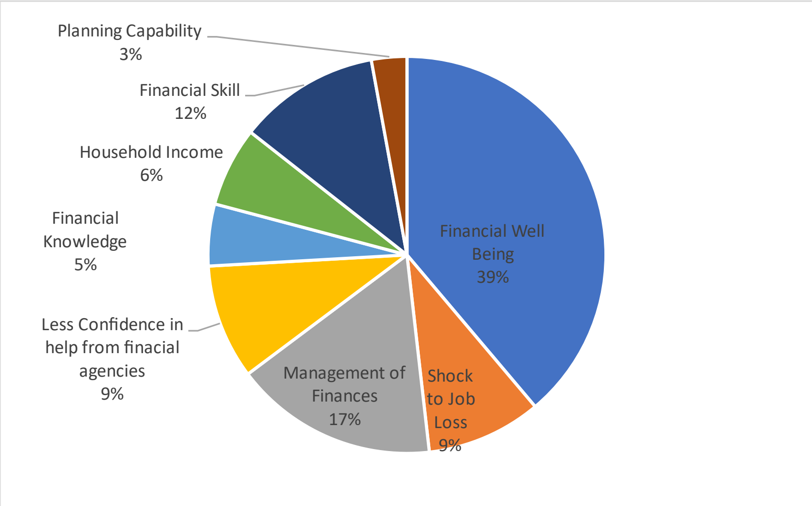


Figure: Category-wise Variable Influence

The influence of Financial Well- being on an individual's chances of material hardship is very significant. By implementing certain policies to maintain a healthy financial well being the risk of this crisis can be evaded. Also, the influence of management of finances is evident which focuses on the financial agencies duties to guide an individual to manage their finances well so that they do not experience the difficulties of hardship.

Financial Skill and Financial knowledge have a crucial impact on an individual's likelihood of encountering the crisis, so, planning of financial literacy programs in areas in a timely manner is an endeavour that can be undertaken for the betterment of the scenario.

As, households or individuals suffer a lot from bad experience with financial agencies these organizations need to be cautious while dealing with people and thus, can identify and aid individuals to avoid the material hardship experiencing scenario.

Besides, there can be tailored policies and programs for betterment of the individuals and society overall.

I. Financial Literacy Programs:

Implementing thorough financial literacy initiatives aimed at disadvantaged groups of people is vital. Improving one's knowledge of finance can allow people to make better decisions, which may lower their chance of facing monetary hardship.

II. Targeted Social Safety Net Interventions:

Social safety net initiatives need to be customized to encounter the unique requirements that predictive models have discovered. Tailored interventions have the potential to efficiently distribute resources, offering rapid assistance to the most vulnerable.

III. Early Warning Systems:

It is important to set up procedures for recurring evaluations using forecasting models. Hands-on interventions can be made to circumvent the escalation of material hardship by early identification of those at risk.

IV. Community Outreach and Education:

To raise awareness of the resources and support services that are available, conducting community outreach initiatives is crucial. Communities with more knowledge are better able to take use of and access the resources already in place.

6.2.1. Future Work

i. Dynamic Model Refinement:

Refining prediction models over time using new data and changing factors need to be considered. Economic environments shift, yet dynamic models maintain their accuracy and relevance throughout time.

ii. Ethical and Bias Considerations:

Conducting rigorous assessments of ethical considerations and biases in predictive models will aid the betterment of the model. Ensuring fairness and transparency is supreme, especially when the models impact vulnerable populations.

iii. Intersectionality Analysis:

Examining the interplay between several elements that lead to financial difficulties will improve the model. Determining the interplay between various components might result in more sophisticated policy suggestions.

7. Conclusion

This thesis offers significant insights into the intricate world of economic vulnerability via investigating material suffering, its causes, and the creation of prediction algorithms meant to detect those who are at risk. The outcome of the logistic regression, support vector machine, Naive Bayes, and decision tree models discovered the vital factors affecting material hardship. Three key predictors surfaced: the Financial Well-Being question, gift-giving concerns, and financial agency experiences. The several models provided complimentary viewpoints, highlighting the phenomenon's complexity.

Although predictive models offer momentous quantitative insights, it is important to comprehend their limits. Priorities must be given to ethical issues, continuous validation, and correcting prejudices. Material hardship is not only a quantitative phenomenon; qualitative studies can add depth to these models by elucidating the complex human dynamics associated with financial hardships. This research contributes to the continuing discussion on economic vulnerability and informs practical tactics for promoting a more resilient and equitable society by fusing academic ideas with data-driven methodologies. As I advance, navigating the complex terrain of economic well-being will continue to need an integrated and multidisciplinary approach.

Reference:

1. Thomas, M.M.C. Longitudinal Patterns of Material Hardship Among US Families. Soc Indic Res 163, 341–370 (2022). https://doi.org/10.1007/s11205-022-02896-8

2. Nelson, G. (2012). Measuring Poverty: The Official U.S. Measure and Material Hardship. Journal of Poverty, 16(3), 245-263. First published on 10th August 2012. DOI: https://doi.org/10.2202/1944-2858.1077

3. Heflin, C. M., & Olson, K. (2017). Unpacking an anomaly: Analyzing the 2005–2012 current population survey to understand why many emergency food assistance users report being food secure. Journal of Hunger & Environmental Nutrition, 12(1), 64–76. https://doi.org/10.1080/19320248.2016.1157548

4. Beverly, S. G. (2008). Measures of Material Hardship: Rationale and Recommendations. Journal of Poverty, 12(4), 389-410. DOI: https://doi.org/10.1300/J134v05n01_02

5. McKernan, S. M., Ratcliffe, C., & Braga, B. (2021). The effect of the US safety net on material hardship over two decades. Journal of Public Economics, 199, 104403. DOI: https://doi.org/10.1016/j.jpubeco.2021.104403

6. Ashiabi, G. S., & O'Neal, K. K. (2007). Children's Health Status: Examining the Associations among Income Poverty, Material Hardship, and Parental Factors. PLoS ONE, 2(9), e940. Published on September 26, 2007. DOI: https://doi.org/10.1371/journal.pone.0000940

7. Chaudry, A., & Wimer, C. (2016). Poverty is Not Just an Indicator: The Relationship Between Income, Poverty, and Child Well-Being. Academic Pediatrics, 16(3 Suppl), S23-S29. PMID: 27044698. DOI: 10.1016/j.acap.2015.12.010

8. Marçal, K. E. (2022). Patterns and predictors of material hardship among poor families with children. Journal of Community Psychology, [Advance online publication]. https://doi.org/10.1002/jcop.22978

9. Duncan, G. J., Ziol-Guest, K. M., & Kalil, A. (2010). Early-childhood poverty and adult attainment, behavior, and health. Child Development, 81(1), 306-325. DOI: 10.1111/j.1467-8624.2009.01396.x

10. Zilanawala, A., & Pilkauskas, N. V. (2012). Material hardship and child socioemotional behaviors: Differences by types of hardship, timing, and duration. Children and Youth Services Review, 34(4), 814–825. https://doi.org/10.1016/j.childyouth.2012.01.008

11. Yoo, J. P., Slack, K. S., & Holl, J. L. (2009). Material hardship and the physical health of school-aged children in low-income households. American Journal of Public Health, 99(5), 829–836. https://doi.org/10.2105/AJPH.2007.119776

12. Beverly, S. (2001). Material Hardship in the United States: Evidence from the Survey of Income and Program Participation. Social Work Research, 25(3), 143-151. from DOI: 10.1093/swr/25.3.143

13. Beverly, S. (2001). Material Hardship in the United States: Evidence from the Survey of Income and Program Participation. Social Work Research, 25(3), 143-151. from DOI: 10.1093/swr/25.3.143

14. Beverly, S. (2001). Material Hardship in the United States: Evidence from the Survey of Income and Program Participation. Social Work Research, 25(3), 143-151. from DOI: 10.1093/swr/25.3.143

15. Beverly, S. (2001). Material Hardship in the United States: Evidence from the Survey of Income and Program Participation. Social Work Research, 25(3), 143-151. from DOI: 10.1093/swr/25.3.143

16. Beverly, S. (2001). Material Hardship in the United States: Evidence from the Survey of Income and Program Participation. Social Work Research, 25(3), 143-151. from DOI: 10.1093/swr/25.3.143

17. Hastings, J., Madrian, B. and Skimmyhorn, W. (2013) Financial Literacy, Financial Education, and Economic Outcomes. Annual Review of Economics, 5, 347-373. https://doi.org/10.1146/annurev-economics-082312-125807

18. https://www.theguardian.com/us-news/2021/oct/12/us-households-financial-problems-medical-care-food

19. Gershoff, E. T., Aber, J. L., Raver, C. C., & Lennon, M. C. (2007). Income Is Not Enough: Incorporating Material Hardship Into Models of Income Associations With Parenting and Child Development. Child Development, 78(1), 70-95. First published on 28th February 2007. DOI: https://doi.org/10.1111/j.1467-8624.2007.00986.x

20. Sullivan, J. X., Turner, L., & Danziger, S. (2007). The relationship between income and material hardship. Published in the Journal of Policy Analysis and Management, on 4th December 2007. DOI: https://doi.org/10.1002/pam.20307. Citations: 121

21. Heflin, C. M. (2017). The role of social positioning in observed patterns of material hardship: New evidence from the 2008 Survey of Income and Program Participation. Social Problems. https://doi.org/10.1093/socpro/spw041

22. Chalfin, Aaron, Oren Danieli, Andrew Hillis, Zubin Jelveh, Michael Luca, Jens Ludwig, and Sendhil Mullainathan. 2016. "Productivity and Selection of Human Capital with Machine Learning." American Economic Review, 106 (5): 124-27. DOI: 10.1257/aer.p20161029

23. Desai, Ajit, Machine Learning for Economics Research: When What and How? (March 27, 2023). https://dx.doi.org/10.2139/ssrn.4404772

24. Athey, Susan. "The Impact of Machine Learning on Economics." The Economics of Artificial Intelligence: An Agenda (forthcoming). University of Chicago Press, January 2018

25. https://www.consumerfinance.gov/data-research/financial-well-being-survey-data/

Appendix

FWB1_1	category
FWB1_2	category
FWB2_1	category
FWB2_2	category
FS1_1	category
FS1_2	category
FS1_3	category
FS1_4	category
FS1_5	category
FS1_6	category
FS1_7	category
FS2_1	category
FS2_2	category
FS2_3	category
ACT1_1	category
ACT1_2	category
FINGOALS	category
PROPPLAN_1	category
PROPPLAN_2	category
PROPPLAN_3	category
PROPPLAN_4	category
MANAGE1_1	category
MANAGE1_2	category
MANAGE1_3	category
MANAGE1_4	category
SAVEHABIT	category
FRUGALITY	category
ASK1_1	category
ASK1_2	category
CHANGEABLE	category
FINKNOWL1	category
FINKNOWL2	category
FINKNOWL3	category
KHKNOWL1	category
KHKNOWL2	category
KHKNOWL3	category
KHKNOWL4	category

Exhibit 1: Category Types of variables

X Training Data	

Number of rows: 3385	
Number of columns: 281	
X Testing Data	

Number of rows: 1452	
Number of columns: 281	
y Training Data	

Number of rows: 3385	
y Testing Data	

Number of rows: 1452	

Exhibit 2: Testing and training data

Logistic Regression	**Decision Tree Classifier**	**Support Vector Classifier**	**Bernoulli Naive Bayes Classifier**
'C': 0.1	'min_samples_split': 40	'C': 1	'alpha': 12
'class_weight': None	‘min_samples_leaf': 2	'class_weight': 'balanced'	
	‘min_impurity_d	'gamma':	

	ecrease': 0.0005	'scale'	
	‘max_features': 'sqrt'	kernel': 'rbf'	
	‘max_depth': 3		
	'criterion': 'entropy'		
	'class_weight': 'balanced'		

Exhibit 3: Best Parameters of models

Logistic Regression Feature Importance	
FWB2_1_5 (Giving a gift...would put a strain on my finances for the month-ALWAYS)	0.6954
FS1_1_1 (I know how to get myself to follow through on my financial intentions-NOT AT ALL)	0.6061
FWB2_2_1 (I have money left over at the end of-NEVER)	0.5903
FWB2_1_1 (Giving a gift...would put a strain on my finances for the month-NEVER)	-0.5865
FWB2_1_4 (Giving a gift...would put a strain on my finances for the month- OFTEN)	0.5391
FWB2_1_2(Giving a gift...would put a strain on my finances for the month- RARELY)	-0.5053
FWB1_1_4 (I could handle a major unexpected expense-VERY WELL)	-0.4883
EMPLOY_7 (Unemployed or temporarily laid off)	0.4848
MANAGE1_1_3 (Paid all your bills on time-SOMETIMES)	0.4821
MANAGE1_1_5 ((Paid all your bills on time-ALWAYS))	-0.4811
EMPLOY_8 (Retired)	-0.4566
REJECTED_2_0 (I did not apply for credit because I thought would be turned down-NO)	-0.4261

REJECTED_2_1(I did not apply for credit because I thought would be turned down-YES)	0.4260
SHOCKS_1_0 (LOST A JOB-NO)	-0.4140
SHOCKS_1_1 (LOST A JOB-YES)	0.4139
FWB2_2_5 (I have money left over at the end of the month- ALWAYS)	-0.3494
PPINCIMP_8 (Household Income $100,000 to $149,999)	-0.3476
FS1_6_3 (I know how to keep myself from spending too much-SOMEWHAT)	0.3402
agecat_3 (AGE 35-44)	0.3169
FS1_2_1(I know where to find the advice I need to make decisions involving money-NOT AT ALL)	0.3150

Exhibit 4: Logistic Regression Feature Importance

Logistic Regression Odds Ratio	
FWB2_1_5(Giving a gift...would put a strain on my finances for the month-ALWAYS)	2.00
FWB1_1_1(I know how to get myself to follow through on my financial intentions-NOT AT ALL)	1.83
FWB2_2_1(I have money left over at the end of the Month- NEVER)	1.80
FWB2_1_4(Giving a gift...would put a strain on my finances for the month- OFTEN)	1.71
EMPLOY_7(Unemployed or temporarily laid off)	1.62
MANAGE1_1_3(Paid all your bills on time-SOMETIMES)	1.62
REJECTED_2_1(I did not apply for credit because I thought would be turned down-YES)	1.53
SHOCKS_1_1(LOST A JOB-YES)	1.51
FS1_6_3(I know how to keep myself from spending too much-SOMEWHAT)	1.41
agecat_3 (AGE 35-44))	1.3

FS2_3_1 (I struggle to understand financial information- NEVER)	0.75
PPINCIMP_8(Household Income $100,000 to $149,999)	0.71
FWB2_2_5(I have money left over at the end of the month- ALWAYS)	0.71
SHOCKS_1_0(LOST A JOB-YES)	0.66
REJECTED_2_0(I did not apply for credit because I thought would be turned down-NO)	0.65
EMPLOY_8(Retired)	0.63
MANAGE1_1_5((Paid all your bills on time- ALWAYS))	0.62
FWB1_1_4(I could handle a major unexpected expense-VERY WELL)	0.61
FWB2_1_2(Giving a gift...would put a strain on my finances for the month- RARELY)	0.60
FWB2_1_1(Giving a gift...would put a strain on my finances for the month- NEVER)	0.56

Exhibit 5: Logistic Regression Odds Ratio

Support Vector Feature Importance	
FWB2_1_5(Giving a gift...would put a strain on my finances for the month-ALWAYS)	32.5652
FWB1_1_1(I know how to get myself to follow through on my financial intentions- NOT AT ALL)	32.0769
FWB2_1_2(Giving a gift...would put a strain on my finances for the month- RARELY)	27.5058
MANAGE1_1_3(Paid all your bills on time- SOMETIMES)	26.7373
FWB2_1_4(Giving a gift...would put a strain on my finances for the month- OFTEN)	26.6296
FWB2_2_1(I have money left over at the end of the Month- NEVER)	25.8231
REJECTED_2_1(I did not apply for credit because I thought would be turned down-YES)	25.0865
REJECTED_2_0(I did not apply for credit because I thought would be turned down-NO)	25.0865

MANAGE1_1_5 ((Paid all your bills on time-ALWAYS))	23.3040
FWB2_1_1 (Giving a gift...would put a strain on my finances for the month-NEVER)	22.5080
FWB1_1_4 (I could handle a major unexpected expense-VERY WELL)	19.8860
EMPLOY_8(Retired)	17.9187
EMPLOY_7 (Unemployed or temporarily laid off)	17.5335
PPINCIMP_1 (Less than $20,000)	15.9850
PPINCIMP_8 (Household Income $100,000 to $149,999)	15.4555
FS1_6_3(I know how to keep myself from spending too much-SOMEWHAT)	14.9397
agecat_2 (25-34)	14.5444
SHOCKS_1_1(LOST A JOB-YES)	14.3804
SHOCKS_1_0 (LOST A JOB-NO)	14.3804
MANAGE1_3_3 (Paid off credit card balance in full each month-SOMETIMES)	14.2142

Exhibit 6: Feature Importance Support Vector

Support Vector Permutation Importance	
MANAGE1_1_5((Paid all your bills on time-ALWAYS))	0.0100
FWB2_1_5(Giving a gift...would put a strain on my finances for the month-ALWAYS)	0.0082
FWB2_2_1(I have money left over at the end of the Month- NEVER)	0.0071
SHOCKS_1_1(LOST A JOB-YES)	0.0045
SHOCKS_1_0(LOST A JOB-NO)	0.0045
FWB2_1_4(Giving a gift...would put a strain on my finances for the month- OFTEN)	0.0044
FWB2_1_2(Giving a gift...would put a strain on my finances for the month- RARELY)	0.0001
agecat_2 (25-34)	0.0027
EMPLOY_8 (Retired)	0.0025

FWB2_1_1(Giving a gift...would put a strain on my finances for the month-NEVER)	0.0026
FS1_6_3(I know how to keep myself from spending too much-SOMEWHAT)	0.0015
PPINCIMP_8(Household Income $100,000 to $149,999)	0.0010
FWB1_1_1(I know how to get myself to follow through on my financial intentions-NOT AT ALL)	0.0004
MANAGE1_1_3(Paid off credit card balance in full each month-SOMETIMES)	-0.0010
REJECTED_2_0(I did not apply for credit because I thought would be turned down-NO)	-0.0032
REJECTED_2_1(I did not apply for credit because I thought would be turned down-YES)	-0.0032
FWB1_1_4(I could handle a major unexpected expense-VERY WELL)	-0.0018
EMPLOY_7 ((Unemployed or temporarily laid off))	-0.0002
PPINCIMP_1(Less than $20,000)	-0.0004
MANAGE1_3_3(Paid off credit card balance in full each month-SOMETIMES)	-0.0024

Exhibit 7: Permutation Importance Support Vector

Naïve Bayes Log Probability Difference	
FWB2_1_5(Giving a gift...would put a strain on my finances for the month-ALWAYS)	2.16
MANAGE1_1_2	2.14
FWB2_2_1(I have money left over at the end of the Month- NEVER)	1.94
SNAP_1(Any household member received SNAP benefits-YES)	1.87
REJECTED_2_1(I did not apply for credit because I thought would be turned down-	1.84

YES)	
MANAGE1_1_3(Paid off credit card balance in full each month-SOMETIMES)	1.83
FS1_7_1 (I know how to make myself save-NOT AT ALL)	1.82
ACT1_1_2 (I follow-through on my financial commitments to others-VERY LITTLE)	1.82
FS1_1_1 (I know how to get myself to follow through on my financial intentions- NOT AT ALL)	1.78
FWB1_2_1 (I am securing my financial future- NOT AT ALL)	1.74
EMPLOY_7(Unemployed or temporarily laid off)	1.71
ACT1_1_1(I follow-through on my financial commitments to others- NOT AT ALL)	1.7
FRUGALITY_1 (If I can re-use an item I already have, there's no sense in buying something new- Strongly disagree)	1.7
ACT1_2_1(I follow-through on financial goals I set for myself- NOT AT ALL)	1.65
FWB1_1_1(I could handle a major unexpected expense- NOT AT ALL)	1.63
SHOCKS_1_1(LOST A JOB-YES)	1.61
SAVEHABIT_1 (Putting money into savings is a habit for me- Strongly disagree)	1.58
FS1_6_1 (I know how to keep myself from spending too much- NOT AT ALL)	1.57
MANAGE1_1_1(Paid all your bills on time-NEVER)	1.56
EMPLOY_6 (Permanently sick, disabled or unable to work**)**	1.55

Exhibit 8: Log Probability Difference of Naïve Bayes model

Naïve Bayes Permutation Importance

SNAP_1(Any household member received SNAP benefits-YES)	0.0018
SNAP_0(Any household member received SNAP benefits-NO)	0.0018
FWB2_1_5(Giving a gift...would put a strain on my finances for the month-ALWAYS)	0.0017
KHKNOWL6_1 (Understanding of possibility of housing market losses-TRUE)	0.0010
KH6correct_1 (KHKNOWL6 answered correctly-YES)	0.0010
KHKNOWL6_2 (Understanding of possibility of housing market losses-FALSE)	0.0010
KH6correct_0 (KHKNOWL6 answered correctly-NO)	0.0010
FWB2_2_1(I have money left over at the end of the Month- NEVER)	0.0006
FS2_3_5 (I struggle to understand financial information- ALWAYS)	0.0006
PPHHSIZE_5 (Household Size- 5+)	0.0006
KHKNOWL7_1 (Understanding of credit card minimum payments- Less than 5 years)	0.0006
FS1_3_1 (I know how to make complex financial decisions- NOT AT ALL**)**	0.0006
FS1_2_1 (I know where to find the advice I need to make decisions involving money- NOT AT ALL)	0.0006
KHKNOWL2_3 (Understanding of stocks vs bond vs savings volatility-STOCKS)	0.0005
KH2correct_1 (KHKNOWL2 answered correctly-YES)	0.0005
KH2correct_0 ((KHKNOWL2 answered correctly-NO)	0.0005
KHKNOWL2_1 (Understanding of stocks vs bond vs savings volatility-SAVINGS ACCOUNTS)	0.0004
FRAUD2_8 (Victim of financial fraud or attempted financial fraud in past 5 years)	0.0004
FRUGALITY_3 (If I can re-use an item I already have, there's no sense in buying something new- Disagree slightly)	0.0004
PPINCIMP_3 (Household Income -$30,000 to $39,999)	0.0003

Exhibit 9: Permutation Importance of Naïve Bayes model

Decision Tree Feature Importance	
FWB1_1_1(I know how to get myself to follow through on my financial intentions- NOT AT ALL)	0.3354
MANAGE1_1_5(Paid off credit card balance in full each month- ALWAYS)	0.1930
REJECTED_2_0(I did not apply for credit because I thought would be turned down-NO)	0.0783
MANAGE1_1_4 ((Paid off credit card balance in full each month- OFTEN**)**	0.0640
FWB1_1_2(I know how to get myself to follow through on my financial intentions- Very little)	0.0540
SNAP_1(Any household member received SNAP benefits-YES)	0.0435
SHOCKS_1_1(LOST A JOB-YES)	0.0220
KHKNOWL6_1(KHKNOWL6 answered correctly-YES)	0.0217
FWB2_1_5(Giving a gift...would put a strain on my finances for the month-ALWAYS)	0.0203
KHKNOWL3_1 (Understanding of benefits of diversification- INCREASE)	0.0169
FWB1_1_3	0.0165
FWB2_1_2(Giving a gift...would put a strain on my finances for the month- RARELY**)**	0.0145
PROPPLAN_4_2	0.0095
KH2correct_0	0.0092
MANAGE1_4_5	0.0081
FWB2_2_4	0.0077
PPGENDER_1	0.0074
EMPLOY_8	0.0071
FRAUD2_0	0.0064
KH1correct_0	0.0062

Exhibit 10: Feature Importance of Decision Tree

```
Logistic Regression Equation: P(Facing Major Hardship=1) = 1 / (1 + e^([-0.0001906] + 0.5230570691616832*FWB1_1_1 + 0.027516432
036587376*FWB1_1_2 + -0.06327938748623811*FWB1_1_3 + -0.4635015851300115*FWB1_1_4 + -0.024189537467740635*FWB1_1_5 + 0.02296930
239189602*FWB1_2_1 + 0.18514148269345757*FWB1_2_2 + 0.11192016889734031*FWB1_2_3 + -0.2823196001949357*FWB1_2_4 + -0.0381083626
73476844*FWB1_2_5 + -0.5180739405115399*FWB2_1_1 + -0.516722128165896 9*FWB2_1_2 + -0.19668306935958405*FWB2_1_3 + 0.55113534040
4016*FWB2_1_4 + 0.679946788747285*FWB2_1_5 + 0.5431123569564109*FWB2_2_1 + 0.07068671490612162*FWB2_2_2 + -0.03058494190707306*
FWB2_2_3 + -0.23310886779340156*FWB2_2_4 + -0.3505022710477767*FWB2_2_5 + 0.030505932728614137*FS1_1_1 + 0.12305954386161365*FS
1_1_2 + 0.00606569768929215*FS1_1_3 + 0.05784301174801755*FS1_1_4 + -0.21787119491325613*FS1_1_5 + 0.28145793932051877*FS1_2_1
+ 0.12653708722168977*FS1_2_2 + -0.04296284650738*FS1_2_3 + -0.1608280545501445*FS1_2_4 + -0.20460113437040214*FS1_2_5 + -0.249
5950077547047*FS1_3_1 + -0.03294287187790775*FS1_3_2 + 0.08017416197546107*FS1_3_3 + 0.08553594400603895*FS1_3_4 + 0.1164307647
6539382*FS1_3_5 + 0.19795073001857097*FS1_4_1 + -0.3526534753164404*FS1_4_2 + 0.01808092258380203*FS1_4_3 + 0.0501293566759557*
FS1_4_4 + 0.08609545715239335*FS1_4_5 + 0.11340613207287743*FS1_5_1 + -0.16844483478034686*FS1_5_2 + -0.1102266955780018 7*FS1_5
_3 + -0.008101395292919762*FS1_5_4 + 0.17296978469267232*FS1_5_5 + -0.064381303635698*FS1_6_1 + -0.1970439432256777*FS1_6_2 +
0.24315241916248315*FS1_6_3 + -0.1817007333455581*FS1_6_4 + 0.19957655215873144*FS1_6_5 + -0.055799046453449075*FS1_7_1 + 0.170
80976562320477*FS1_7_2 + -0.01927536771650663*FS1_7_3 + -0.02679130680537525*FS1_7_4 + -0.0693410535335924*FS1_7_5 + -0.0465148
74485082215*FS2_1_1 + -0.0566830023105964*FS2_1_2 + 0.005509021509434855*FS2_1_3 + -0.06152607392779 6895*FS2_1_4 + 0.1588179203
2832204*FS2_1_5 + 0.05804111684246366*FS2_2_1 + 0.11105243685728075*FS2_2_2 + -0.03523044067422541*FS2_2_3 + -0.192128745711916
77*FS2_2_4 + 0.05786862380067901*FS2_2_5 + -0.14927689987964418*FS2_3_1 + -0.0621342031077809*FS2_3_2 + 0.021803274680935986*FS
2_3_3 + 0.02502172809642216*FS2_3_4 + 0.16418909132434806*FS2_3_5 + 0.04692531442899531*ACT1_1_1 + 0.13370806672728083*ACT1_1_2
+ 0.013868954816264656*ACT1_1_3 + 0.03927461010686019*ACT1_1_4 + -0.2341739549651195*ACT1_1_5 + -0.08088214401454896*ACT1_2_1 +
0.050595790726130706*ACT1_2_2 + 0.10758419272154937*ACT1_2_3 + -0.04067376615607425*ACT1_2_4 + -0.03702108216277533*ACT1_2_5 +
0.06371349502616623*FINGOALS_0 + -0.06411050391188494*FINGOALS_1 + 0.011477310368170488*PROPPLAN_1_1 + -0.14536112733276388*PRO
PPLAN_1_2 + 0.15061287397398992*PROPPLAN_1_3 + -0.09470024871570745*PROPPLAN_1_4 + 0.0775741828205922*PROPPLAN_1_5 + 0.05024024
972466749*PROPPLAN_2_1 + -0.2155843720534103 3*PROPPLAN_2_2 + 0.03197611705590343*PROPPLAN_2_3 + 0.05127673307035517*PROPPLAN_2_
4 + 0.08169426331676562*PROPPLAN_2_5 + -0.10359261945748115*PROPPLAN_3_1 + -0.24576854378653865*PROPPLAN_3_2 + 0.01133390893563
1609*PROPPLAN_3_3 + 0.2175017514326268*PROPPLAN_3_4 + 0.12012849399004252*PROPPLAN_3_5 + 0.15246697320525468*PROPPLAN_4_1 + -0.
08602623536255057*PROPPLAN_4_2 + 0.020232504013814566*PROPPLAN_4_3 + 0.11862642130971864*PROPPLAN_4_4 + -0.20569667205195577*PR
OPPLAN_4_5 + -0.1015413274619195*MANAGE1_1_1 + 0.22248042007338442*MANAGE1_1_2 + 0.33927594962550683*MANAGE1_1_3 + -0.033428765
85684707*MANAGE1_1_4 + -0.4271832852658433*MANAGE1_1_5 + -0.13257849594472446*MANAGE1_2_1 + 0.24620281026843552*MANAGE1_2_2 + -
0.04011510737095085*MANAGE1_2_3 + -0.11360585418904963*MANAGE1_2_4 + 0.03969963835057037*MANAGE1_2_5 + 0.18053323859841064*MANA
GE1_3_1 + 0.19118802017377093*MANAGE1_3_2 + 0.05753851147379837*MANAGE1_3_3 + -0.1533121962071579*MANAGE1_3_4 + -0.276344582924
54084*MANAGE1_3_5 + -0.1105429643061734*MANAGE1_4_1 + -0.04065548019822429*MANAGE1_4_2 + 0.20504086283489842*MANAGE1_4_3 + -0.0
5715452569146471*MANAGE1_4_4 + 0.002915098475245 2855*MANAGE1_4_5 + 0.14427361338657324*SAVEHABIT_1 + 0.20148049750160282*SAVEHA
BIT_2 + -0.1691346478066001*SAVEHABIT_3 + -0.09245817305606495*SAVEHABIT_4 + -0.10970799480342738*SAVEHABIT_5 + 0.0251496958921
97524*SAVEHABIT_6 + -0.05432626808812600 3*FRUGALITY_1 + -0.17384904816585386*FRUGALITY_2 + 0.12929808036462126*FRUGALITY_3 + 0.
08322031469621845*FRUGALITY_4 + -0.12913600115817675*FRUGALITY_5 + 0.144395913465597 8*FRUGALITY_6 + -0.23912640430145 51*ASK1_1_
1 + -0.07713774380496079*ASK1_1_2 + -0.025720017305385357*ASK1_1_3 + 0.3496528948835625*ASK1_1_4 + -0.008065738357480005*ASK1_1
_5 + -0.2130148541773701*ASK1_2_1 + -0.10413746957406554*ASK1_2_2 + 0.18137657715218522*ASK1_2_3 + 0.10823984667099483*ASK1_2_4
+ 0.027138891042537017*ASK1_2_5 + 0.2267846325046181*CHANGEABLE_1 + -0.13451743886519738*CHANGEABLE_2 + -0.156681266021800 8*CHA
NGEABLE_3 + -0.16121107151110242*CHANGEABLE_4 + 0.026882615152038554*CHANGEABLE_5 + -0.11362470211314023*CHANGEABLE_6 + 0.31197
022196886554*CHANGEABLE_7 + -0.006821377092297477*FINKNOWL1_1 + -0.12750300699415934*FINKNOWL1_2 + 0.13392737520073775*FINKNOWL
1_3 + -0.09048861427798804*FINKNOWL2_1 + 0.13316129566592028*FINKNOWL2_2 + -0.04306969027365093*FINKNOWL2_3 + -0.02777095950259
8884*FINKNOWL3_1 + 0.027373950616880013*FINKNOWL3_2 + 0.20596797199349584*KHKNOWL1_1 + -0.16064645688809917*KHKNOWL1_2 + -0.045
71852399111504*KHKNOWL1_3 + 0.07597744231130492*KHKNOWL2_1 + 0.08625421464138391*KHKNOWL2_2 + -0.16262866583840732*KHKNOWL2_3 +
0.08668361794285717*KHKNOWL3_1 + -0.06580203212733635*KHKNOWL3_2 + -0.021278594701239224*KHKNOWL3_3 + 0.039555456381160044*KHKN
OWL4_1 + -0.039952465266878845*KHKNOWL4_2 + -0.05605748431690446*KHKNOWL5_1 + 0.055660475431185914*KHKNOWL5_2 + 0.1438422903778
2637*KHKNOWL6_1 + -0.1442392992635453*KHKNOWL6_2 + 0.13929820108857977*KHKNOWL7_1 + 0.1066779518545568*KHKNOWL7_2 + -0.18056533
038262104*KHKNOWL7_3 + -0.06580783144623414*KHKNOWL7_4 + 0.0447783000408619*KHKNOWL8_1 + -0.01645310304278539*KHKNOWL8_2 + 0.01
8461335023545418*KHKNOWL8_3 + -0.04718354090734093 5*KHKNOWL8_4 + -0.06573655762040333*KHKNOWL9_1 + 0.06533954873468459*KHKNOWL9
_2 + 0.045321515105396164*KH1correct_0 + -0.04571852399111504*KH1correct_1 + 0.16223165695268885*KH2correct_0 + -0.162628665838

018599*PPINCIMP_1 + 0.21361232508695321*PPINCIMP_2 + 0.19584460212323523*PPINCIMP_3 + 0.15791861674291333*PPINCIMP_4 + 0.023790
47242360298*PPINCIMP_5 + 0.04953062016116661*PPINCIMP_6 + -0.2719086230965648*PPINCIMP_7 + -0.4960380041107938*PPINCIMP_8 + -0.
1490987570180915*PPINCIMP_9 + 0.13584010110745504*agecat_1 + 0.33222547307644995*agecat_2 + 0.39586980878343 35*agecat_3 + 0.116
19377208993907*agecat_4 + -0.10095875656570075*agecat_5 + -0.3569208263862768*agecat_6 + -0.27672956104443874*agecat_7 + -0.245
91701994657963*agecat_8 + -0.05282225028369376 4*KIDS_NoChildren_0 + 0.052425241397975296*KIDS_NoChildren_1 + 0.021082213146290 6
66*SELFCONTROL_3_1 + 0.15279586920356092*SELFCONTROL_3_2 + -0.013713190350212985*SELFCONTROL_3_3 + -0.1605619008853576 3*SELFCON
TROL_3_4 + -0.15215101576310414*HHEDUC_1 + -0.037492426321223744*HHEDUC_2 + 0.2321687493295115*HHEDUC_3 + 0.002878774079617832*
HHEDUC_4 + -0.0458010902105199 6*HHEDUC_5 + 0.006424368206578323*FK1correct_0 + -0.006821377092297477*FK1correct_1 + 0.042672681
38793225*FK2correct_0 + -0.04306969027365093*FK2correct_1 + -0.027770959502598884*FK3correct_0 + 0.027373950616880013*FK3correc
t_1 + 0.025828150837364 27*SUBNUMERACY1_1 + 0.11342681339187906*SUBNUMERACY1_2 + 0.09204226099411292*SUBNUMERACY1_3 + 0.11446281
964737076*SUBNUMERACY1_4 + -0.20142067872575734*SUBNUMERACY1_5 + -0.14473637503068876*SUBNUMERACY1_6 + -0.18709531171603272*FRA
UD2_0 + 0.1848682120385997*FRAUD2_1 + 0.0018300907917139682*FRAUD2_8 + -0.6099650073937298*CONSPROTECT1_1 + -0.2569897141320499
6*CONSPROTECT1_2 + 0.25752852444649793*CONSPROTECT1_3 + 0.6090291881935632*CONSPROTECT1_4 + -0.01477192099233448 5*CONSPROTECT2_
1 + 0.04577015784345277*CONSPROTECT2_2 + -0.03139524573683734*CONSPROTECT2_3 + -0.30922794507746376*CONSPROTECT3_0 + 0.30883093
61917454*CONSPROTECT3_1 + -0.36559265214722547*REJECTED_2_0 + 0.36519564326150705*REJECTED_2_1 + -0.40970124089651766*SHOCKS_1_
0 + 0.4093042320107983*SHOCKS_1_1 + -0.028060024443882364*HSLOC_1 + 0.027663015558163605*HSLOC_2 + -0.30075484688387966*SNAP_0
+ 0.3003578379981612*SNAP_1))
```

Exhibit 11: Logistic Regression Final Equation

MATHARDSHIP_1	Worried whether food would run out before got money to buy more	-1	Refused	All	27	0.4%
MATHARDSHIP_1	Worried whether food would run out before got money to buy more	1	Never	All	5,224	81.7%
MATHARDSHIP_1	Worried whether food would run out before got money to buy more	2	Sometimes	All	854	13.4%
MATHARDSHIP_1	Worried whether food would run out before got money to buy more	3	Often	All	289	4.5%
MATHARDSHIP_2	Food didn't last and didn't have money to get more	-1	Refused	All	27	0.4%
MATHARDSHIP_2	Food didn't last and didn't have money to get more	1	Never	All	5,360	83.8%
MATHARDSHIP_2	Food didn't last and didn't have money to get more	2	Sometimes	All	767	12.0%
MATHARDSHIP_2	Food didn't last and didn't have money to get more	3	Often	All	240	3.8%
MATHARDSHIP_3	Couldn't afford a place to live	-1	Refused	All	26	0.4%
MATHARDSHIP_3	Couldn't afford a place to live	1	Never	All	5,790	90.6%
MATHARDSHIP_3	Couldn't afford a place to live	2	Sometimes	All	412	6.4%
MATHARDSHIP_3	Couldn't afford a place to live	3	Often	All	166	2.6%
MATHARDSHIP_4	Any household member couldn't afford to see doctor or go to hospital	-1	Refused	All	27	0.4%
MATHARDSHIP_4	Any household member couldn't afford to see doctor or go to hospital	1	Never	All	5,331	83.4%
MATHARDSHIP_4	Any household member couldn't afford to see doctor or go to hospital	2	Sometimes	All	804	12.6%
MATHARDSHIP_4	Any household member couldn't afford to see doctor or go to hospital	3	Often	All	232	3.6%
MATHARDSHIP_5	Any household member stopped taking medication or took less due to costs	-1	Refused	All	25	0.4%
MATHARDSHIP_5	Any household member stopped taking medication or took less due to costs	1	Never	All	5,436	85.0%
MATHARDSHIP_5	Any household member stopped taking medication or took less due to costs	2	Sometimes	All	753	11.8%
MATHARDSHIP_5	Any household member stopped taking medication or took less due to costs	3	Often	All	180	2.8%
MATHARDSHIP_6	Utilities shut off due to non-payment	-1	Refused	All	28	0.4%
MATHARDSHIP_6	Utilities shut off due to non-payment	1	Never	All	5,930	92.7%
MATHARDSHIP_6	Utilities shut off due to non-payment	2	Sometimes	All	356	5.6%
MATHARDSHIP_6	Utilities shut off due to non-payment	3	Often	All	80	1.3%

Exhibit 12: Dependent Variable Queries