Group 15

Advanced Data Analysis Project

Introduction:

The main goal of this project is to analyze and explore the factors influencing the change in the self-perception and goals of youth at risk throughout different periods of life. This analysis can provide insights into the experiences and challenges faced by youth at risk and help develop interventions or support systems to positively influence their self-perception and goals. Those interventions and support systems could promote positive development and improve youth at risk overall well-being.

Prior research has shown that a variety of factors can influence the self-perception and goals of youth. For instance, a study found that adolescents who have strong peer relationships are more likely to have positive self-perceptions and goals ("Peer Relationships and Self-Perceptions in Adolescents: A Study by the University of California, Los Angeles" by Tali Klima and Rena L. Repetti). Other research showed that adolescents with high self-esteem are more likely to have positive self-perceptions and goals ("Self-Esteem and Adolescent Development: A Study by the University of Pennsylvania" by Sheri Feldman).

The main challenge of this research is the fact that there are many different interacting factors at play. Additionally, self-perception and goals are subjective and can be influenced by a variety of social, psychological, and environmental factors. Analyzing and interpreting these complex dynamics requires careful consideration and robust data analysis. This project can extend existing knowledge by conducting a more comprehensive analysis that examines factors over different periods of life, allowing for a deeper understanding of how youth at risk interact and change over time.

Our approach this study is the to use a feature importance model called elastic regression (which combines 2 regression models, lasso and ridge regression) on different questions answered by the youth regarding their self-perception and goals and find the most influential factors according to their answers.

Overview:

The original dataset contained information on 161 teenagers that took the survey at the exact two different life periods that went through adversity during their childhood, there are 535 columns that document what they went through in 3 different periods of their youth such as emotional state, how they feel about their own chances in life, life satisfaction and more.

Out of 535 columns available we chose 50 columns to perform our analysis, there are a lot of questions that are irrelevant to our research which we've elected to pass up on since they won't help our cause. Moreover, there are a lot of questions which have similar context, we tried to pick question which we thought will be the best for our study out of those question that are quite similar.

The columns chosen for this project can be grouped to three major groups:
1. This feature family includes variables such as ID, gender, marital status, number of children, education level, and information related to their military and civil service.

2. self-perception and goals of second period of life: This feature family comprises variables that reflect the self-perception and goals of the children during the second period. It includes measures such as life skills, optimism, future expectations, and life satisfaction.

3. Self-perception and goals of the third period: This feature family encompasses variables representing the self-perception and goals of the children during the third period. It includes measures related to life satisfaction, future expectations, and skills.

Method And Results:

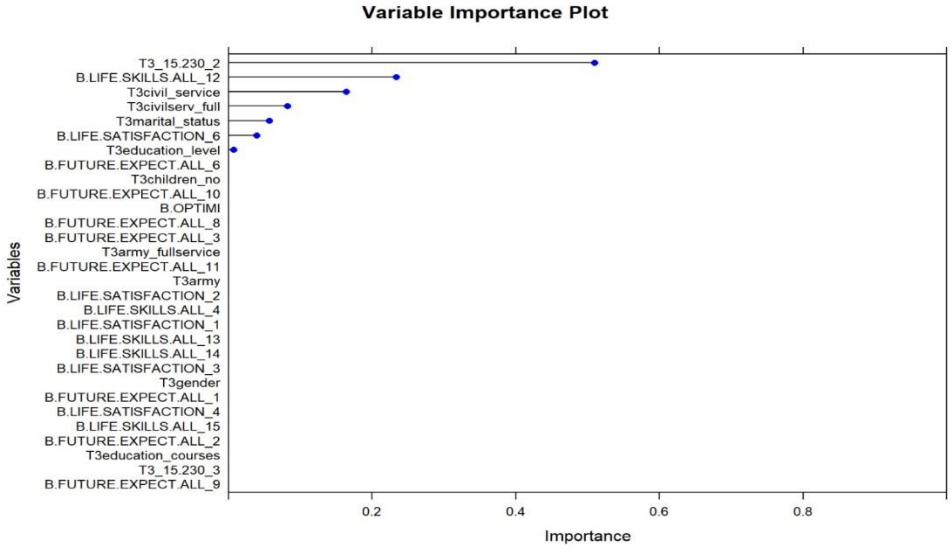
We chose the 'glmnet' package for the data analysis in this project. This model is well-suited for handling a large number of variables and implements Elastic Net regularization, which combines the advantages of lasso regression and ridge regression. By using this model, we aimed to identify the key factors influencing the change in self-perception and goals (feature importance).

The 'glmnet' package is effective in determining the most important predictors in a regression model. It performs feature selection by shrinking the coefficients of irrelevant or redundant predictors towards zero, thus identifying a subset of variables with significant predictive power. This approach enables us to isolate the factors that have the most influence on the desired outcomes.

Through our analysis using the glmnet model, we observed changes in the average scores of various questions over different time periods. We identified a group of factors that consistently exhibited significant influence throughout the different periods, including variables such as depression and being capable of avoiding bad company, another factor which had impact but not as significant as the previous factors is marital status. These factors emerged as the most influential in understanding the changes in self-perception and goals.

To illustrate our findings, let's consider the regression conducted on the question 'Is your life as it should be?' The analysis revealed that the most influential variable was the response to the question 'Have you felt depressed in the last year?'.

Based on the outcome and considering the change in the average scores between the second and third periods which is a decrease of 0.50, we can infer that emotional well-being played a significant role in the decline of life satisfaction levels (see the plot).



In addition to the regression analysis, we conducted result analysis using metrics such as Mean Squared Error (MSE) and Mean Absolute Error (MAE). The MSE value was 5.409247e-07, indicating a low level of error in predicting the target variable. The MAE value was 0.0004943356, reflecting the average absolute difference between the predicted and actual values. We can see similar results at the rest of models we've conducted in our code implementation. These metrics further validate the accuracy and reliability of our model.

Considering the observed changes in the average responses and the insights gained from the analysis, we can infer that the events and factors that transpired between the tests, such as experiencing depression, dealing with personal challenges, and establishing relationships, all had a significant impact on the self-perception and goals of the individuals involved.

Limitation And Future Works:

Firstly, our analysis identified some factors such as: depression, the ability to deal with problems, and marital status as the most influential factors in the change of self-perception and goals among at-risk youth. However, we were unable to determine whether the influence of each factor was positive or negative.

Secondly, it is important to note that our analysis relied on the available data and documented factors. There is a possibility that there are other important variables that were not included in our dataset or were not documented. Future research could focus on exploring additional factors that might contribute to the changes observed in self-perception and goals among at-risk youth. Conducting comprehensive surveys or interviews with the target population in addition to the ones which are documented in the dataset could help uncover such factors and provide a more comprehensive picture.

As for what could be conducted in the future for us, we thought that there is a need to have better understanding of the influence of factors found with high significance, for instance we would like to understand if the military service have a positive or negative influence on self-perception and goals, did it added more pressure or caused more trauma, or maybe it helped the youth and gave tools to deal with some of their problem.

Another intriguing idea was to perform a follow up data analysis is building predictive models using machine learning algorithms to predict the change in self-perception and goals among at-risk youth based on the identified influential factors. This would allow us to forecast future changes and potentially identify high-risk individuals who may require targeted interventions.

Code repository:

https://github.com/RotemEZ/Team-15-Project/tree/main