Predicting Suicidal Thoughts Using Logistic Regression

Felix

2025-03-19

Introduction

Suicidal thoughts are influenced by mental health conditions and economic factors. This report uses **Logistic Regression** with **cross-validation** to predict suicidal thoughts based on mental health and economic indicators.

Load Required Libraries

```
library(caret); library(dplyr); library(readr); library(tidyr)
```

```
## Loading required package: ggplot2

## Loading required package: lattice

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

Load and Preprocess Data

```
mental_health_data <- read.csv("C:/Users/felix/Desktop/CODING/felix's works/Mental-Health-On-Suicide-Ra
economic_data <- read.csv("C:/Users/felix/Desktop/CODING/felix's works/Mental-Health-On-Suicide-Rates-T

mental_health_indicators <- c("Major depressive episode, life", "Eating disorder, current diagnosed con
                              "Social phobia, life", "Generalized anxiety disorder, life", "Bipolar dis
target <- "Suicidal thoughts, life"

mental_health_filtered <- mental_health_data %>%
  filter(Indicators %in% c(mental_health_indicators, target)) %>%
```

```
  select(Year, Geography, Age_Group, Gender, Indicators, Percentage)

mental_health_wide <- mental_health_filtered %>%
  pivot_wider(names_from = Indicators, values_from = Percentage) %>%
  drop_na()

merged_data <- merge(mental_health_wide, economic_data, by.x = c("Year", "Geography"), by.y = c("Year",

for (col in c("CPI", "Inflation_rate")) {
  merged_data[[col]][is.na(merged_data[[col]])] <- median(merged_data[[col]], na.rm = TRUE)
}

features <- c(mental_health_indicators, "CPI", "Inflation_rate")
X <- merged_data[, features]
y <- ifelse(merged_data[[target]] > median(merged_data[[target]], na.rm = TRUE), 1, 0)

X_scaled <- scale(X)
```

Train Logistic Regression Model

```
cv_folds <- trainControl(method = "cv", number = 5)
logistic_model <- train(as.factor(y) ~ ., data = data.frame(X_scaled, y = as.factor(y)),
                        method = "glm", family = "binomial", trControl = cv_folds)
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```
## Warning in nominalTrainWorkflow(x = x, y = y, wts = weights, info = trainInfo,
## : There were missing values in resampled performance measures.
```

```
cat("Cross-Validation Accuracy:", mean(logistic_model$results$Accuracy))
```

```
## Cross-Validation Accuracy: 0.8
```

Conclusion

Logistic regression achieves **80% accuracy** in predicting suicidal thoughts. Future improvements could include **additional economic factors**, **Random Forest models**, or **Neural Networks** for enhanced prediction.