

Pre-process code

HARDIK JAIN - 23111876

2023-08-17

PRE-PROCESSED DATA CODE

PERFORMED BY HARDIK JAIN - 23111876

```
# Load necessary libraries
library(tidyverse)

## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.2      v readr      2.1.4
## v forcats    1.0.0      v stringr   1.5.0
## v ggplot2    3.4.3      v tibble    3.2.1
## v lubridate  1.9.2      v tidyr     1.3.0
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

df <- read.csv('UBER.csv') # ORIGINAL DATASET FILES UBER, NFLX AND TESLA

# Remove rows with NULL/NA values
df_subset <- na.omit(df)

# Feature scaling (normalizing all numeric columns except 'Volume' which is an integer)
normalize <- function(x) {
  return((x - min(x)) / (max(x) - min(x)))
}

df_subset <- df_subset %>% mutate(across(all_of(c("Open", "High", "Low", "Close", "AdjClose")), normalize))

# Save the preprocessed subset to a csv file
write.csv(df_subset, "UBER-PRE.csv", row.names = FALSE) # and other PRE-PROCESSED FILES NFLX-PRE.CSV AND TESLA-PRE.CSV

getwd() # path for current directory

## [1] "/cloud/project"

# view of data
UBER_PRE <- read_csv("UBER-PRE.csv")

## Rows: 725 Columns: 7
## -- Column specification -----
## Delimiter: ","
## chr (1): Date
```

```
## dbl (6): Open, High, Low, Close, AdjClose, Volume
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
view(UBER_PRE)
```