LOW LEVEL DESIGN AND IMPLEMENTATION DOCUMENT

Twitter Sentiment Analysis for Bitcoin Price prediction

UE19CS390B – Capstone Project Phase – 2

Submitted by:

Achyut Jagini	PES2UG19CS013
Kaushal Mahajan	PES2UG19CS178
Namita Aluvathingal	PES2UG19CS245
Vedanth Mohan	PES2UG19CS449

Under the guidance of

Prof. Prajwala
Assistant Professor
PES University

August - December 2022

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
FACULTY OF ENGINEERING
PES UNIVERSITY
(Established under Karnataka Act No. 16 of 2013)
Electronic City, Bengaluru – 560 100, Karnataka, India

PES UNIVERSITY

LOW LEVEL DESIGN AND IMPLEMENTATION DOCUMENT

TABLE OF CONTENTS

1. Introduction

1.1. Overview

In this document, we describe the low-level design and implementation details for the chosen problem statement.

1.2. Purpose

This document provides the details about methodology, approach and implementation of the project.

1.3. Scope

This document shows the approach, architecture, implementation and design of the project.

2. Proposed Methodology / Approach

2.1 Algorithm and Pseudocode

1. Cleaning tweets dataset

- Dropping columns “is_retweet” , “hashtags” , ”source” , ”user_favourites” , “user_friends”.
- Remove rows where all values are NAN.
- Delete rows where “user_verified” value is neither True or False.
- Reduce UTC timestamp to only date value.
- Save dataset as csv

2. Create dataframe of slang data.

- Create a dictionary with abbreviation as the key and the meaning as value using BeatifulSoup.
- Save it as a csv

3.Tweet Preprocessing

- Remove all hashtags , mentions and links from a tweet.
- Lemmatize the text in the tweet.
- Replace slangs in text with their full forms using the slang data
- Find polarity and get sentiment score of tweet using VADER.
- Append a column with the containing the sentiment score for each tweet.
-
- Replace slangs in text with their full forms
- Find polarity scores of text for not lemmatized without slang, not lemmatized with slang ,lemmatized without slang and lemmatized with slang using VADER.
- Lemmatized with slang scores as score in dataset.

4.Extract professions of users

- Using python module get professions of verified users.

- Add professions of the user as a column in the dataset.
- Using the users profession and a list of bitcoin related professions, generate a profession score for each user.
- Append a column to the dataset containing each users' profession score.

5.Get Bitcoin price data

- Create a dataset containing the date, open price, closing price, high and low for the range of dates of the tweets.
- Save this as a csv.

6.Combine datasets

- Normalize user followers and profession score columns of dataset.
- Obtain final score of tweet using sentiment score ,the users' followers and profession score.
- Normalize final score.
- Create a column with the normailised final scores for each tweet.
- Calculate tweet aggregates dataframe with date,tweet volume and average score for each day.
- Merge BTC price data with tweet data using the date as the key.

2.2 Implementation and Results

Implementation

Related professions to calculate profession score

"Financial Analyst" , "Journalist" , "Research Analyst" , "Investment Analyst" , "Cryptocurrency Analyst" , "Blockchain security architect" , "crypto security architect" , "Blockchain Developer" , "Mining technician" , "Consultant", " Trader" , "Software Engineer".

Final score calculation
‘(s["score"] * ((s["user_followers"])) *(s["profession_score"]+1))
score=(tweet_sentiment_score)*(normalised_user_followers)*(profession_score+1)

Slang expansions and abbreviations
Obtained from websites parsed using BeautifulSoup to create slang dictionary.

Results

- Tweet sentiment scores
- User professions
- Tweet profession scores
- Tweet final scores
- Tweet volume
- Average daily tweet score
- Combined dataset

Appendix A: Definitions, Acronyms and Abbreviations

BTC -Bitcoin

VADER- Valence Aware Dictionary and Sentiment Reasoner is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media.

UTC - Coordinated Universal Time

Appendix B: References

- Tweet Sentiment Analysis for Cryptocurrencies"-E. Şaşmaz and F. B. Tek 2021
- Twitter Sentiment Analysis for Bitcoin Price Prediction - Sara Abdali Ben Hoskins
- "Recurrent Neural Network Based Bitcoin Price Prediction by Twitter Sentiment Analysis," -D. R. Pant, P. Neupane, A. Poudel, A. K. Pokhrel and B. K. Lama
- Forecasting Bitcoin Price Fluctuation by Twitter Sentiment Analysis - Otabek Sattarov, Heung Seok Jeon, Ryumduck Oh, Jun Dong Lee
- Cryptocurrency Price Prediction using Sentiment Analysis - Abdul Rehman Khurshid

Appendix C: Record of Change History

#	Date	Document Version No.	Change Description	Reason for Change
1.	12/10/22	1	Added Methodology	
2.	14/10/22	2	Added Results	
3.				