XRP Price Prediction using NLP and LSTM Network Model

[Document subtitle]

Jayden Dzierbicki

2023

Contents

[Git-Repository 2](#_Toc132395002)

[Overview 2](#_Toc132395003)

[WebCrawler 2](#_Toc132395004)

[Data Wrangling 2](#_Toc132395005)

[Machine Learning 2](#_Toc132395006)

[Conclusion & Lessons Learned 2](#_Toc132395007)

# Git-Repository

Documentation and code can be located in a Git-repository:

<https://github.com/jaydendzierbicki/webcrawl-long-short-term-memory-network>

# Overview

The rise of cryptocurrency and blockchain technology has piqued the interest of investors worldwide. Cryptocurrency is a digital currency that uses strong cryptography to verify and protect financial transactions. Ripple Labs is a technology company that developed the Ripple payment protocol and exchange network with a focus on crypto solutions for business and central banks, utilizing their XRP coin. However, XRP, like any other cryptocurrency, experiences significant price fluctuations, posing risks and uncertainties for investors of all levels. To aid investors in making informed decisions, market movement prediction systems have emerged. Traditional supervised learning algorithms have been used in the past to predict changes in cryptocurrency based on historical price data, but predicting price changes can be challenging due to the efficient market hypothesis, which states that the market always follows a haphazard pattern. As a result, investors often actively monitor discussion boards to try to elicit insights that could point to an incoming market movement. However, this task can be overwhelming, and data science can be utilized to develop an automated tool to assist investors. To tackle this challenge, we propose a model which employs natural language processing (NLP) sentiment analysis to forecast price movements within a specific timeframe using the long short-term memory (LSTM) model. The focus is on XRP, owing to its vibrant community, significant market presence, and ongoing developments in the SEC Vs Ripple case in the United States, which has gained widespread media attention. The model aims to encompass novice, intermediate, and experienced investors by collecting data from an array of sources, such as Yahoo Finance and Investing.com, using web crawling and scraping techniques. The data is preprocessed and converted to prepare it for the LSTM network, utilizing Word2Vec to allow the algorithm to comprehend the context of words within the text corpus. However, ethical and legal considerations must be taken into account when gathering and scraping data. ToS on websites form a contractual obligation, and personal information such as usernames or names of posters must not be scraped. By undertaking data gathering, preprocessing, and model development, the proposed model aspires to predict XRP price movements using sentiment analysis and LSTM networks, delivering valuable insights into cryptocurrency price fluctuations.

# WebCrawler

Web crawling and scraping are terms that are often used interchangeably, but they represent distinct concepts and serve unique purposes. When combined, they can be powerful tools for data extraction and analysis. Web scraping involves extracting data from a specific website, where the target site is known, and the data is obtained by inspecting the HTML elements of the webpage (Perez, 2023). On the other hand, web crawling refers to the automated process of navigating the internet to download or index content from multiple websites or URLs. Many sites implement technologies, such as Cloudflare to prevent unauthorized crawling (Gillis, 2022). It is essential to review a website's terms of service before engaging in web crawling or scraping activities, as some sites may have specific rules or restrictions in place which could hinder the automation.

## Methodology

For each domain, we applied a comparable approach by implementing a straightforward extract-transform-load (ETL) procedure. Data was extracted from a designated website, and the extracted data underwent minor transformations, such as adding an ID and minor conversions like separating datetime into date and time components prior to being loaded into a designated table in mySQL. One significant advantage of this method, especially for domains with large datasets, is the enhanced capability for efficient data storage and management. Utilizing a database like MySQL Workbench allows for better scalability, as it can easily accommodate growing data volumes when adding more information to the model. Furthermore, this approach offers seamless integration with Python for further analysis. Through leveraging the storage benefits and scalability offered by a database, we can ensure that our data infrastructure is robust and prepared to handle future growth should we wish to expand our model or add new domains acting as a central repository.

## Domain and workflow

The purpose of this section is to extract data from three different sources as these will later be used as an input layer in our proposed neural network, with two main types of input layers. One input layer will comprise of online forums posts which will later be merged, whilst the second input layer will comprise of financial time series representing the price of XRP-AUD. The reason for obtaining forum posts from two sources was to attempt to capture a variety of different users across different areas of the internet to avoid biases and attempt to capture a larger user base to attempt to include more novice, intermediate and experienced investors. If time persisted, we would suggest scrapping data from more sources.

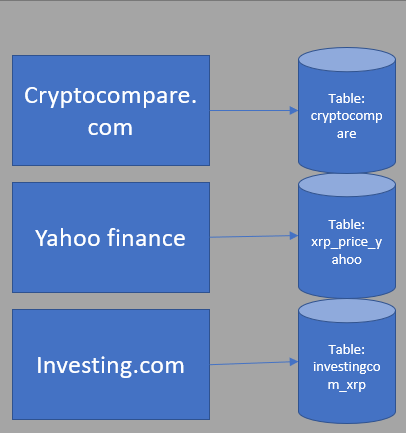


Figure : Utisling the 1-webcrawller.isbp file we were able to crawl and scrape data from various sources, this data underwent an ETL process and was loaded into MySQL tables for data storage.

### Cryptocompare

For us to aqquire

### Yahoo Finance

To acquire the historical price data of XRP-AUD, which is essential for our model, we utilized the **yahoo\_fin** package in Python. By executing the command **get\_data("XRP-AUD", start\_date="01/01/2017", end\_date="15/04/2023", index\_as\_date = False, interval="1d")** in Python, we successfully retrieved daily XRP-AUD data ranging from November 10, 2017 to April 14, 2023. This data contained information on date, opening price, highest price, lowest price, closing price, adjusted closing price, and volume. We then processed the data and imported it into MySQL. This serves as a vital component, as it represents the actual price for a specific day in our machine learning model. It is important to note that the specific time zone employed is not explicitly stated; however, we hypothesize that the data is reported in UTC, as suggested by the XRP price page (XRP AUD (XRP-AUD) price history; historical data 2023).

# Data Wrangling

# Machine Learning

# Conclusion & Lessons Learned

# Reference

G

Gillis, A.S. (2022) What is a web crawler? everything you need to know from techtarget.com, WhatIs.com. TechTarget. Available at: https://www.techtarget.com/whatis/definition/crawler (Accessed: April 18, 2023).

P

Perez, M. (2023) What is web scraping and what is it used for?, ParseHub. Web Scraping Blog (Tips, Guides + Tutorials) | ParseHub. Available at: https://www.parsehub.com/blog/what-is-web-scraping/ (Accessed: April 18, 2023).

X

XRP AUD (XRP-AUD) price history &amp; historical data (2023) Yahoo! Finance. Yahoo! Available at: https://finance.yahoo.com/quote/XRP-AUD/history?p=XRP-AUD (Accessed: April 18, 2023).