NATIONAL UNIVERSITY OF SINGAPORE
IS5006 – INTELLIGENT SYSTEMS DEPLOYMENT
(Semester 2 – AY2020/2021)

Group 3

Final Group Project Title: Intelligent Multi Agent System with Case Based Reasoning

Students Name:

Ngan Zisheng Nigel

Leong Zhong Wei Nicholas

Jeremy Tan

Tan Hui Min

Ang Peng Chuan Alvin

Tan Ming Yang Clarence

Contents

Overview 3

Our PullingAgent collects 3 types of data 4

Price data 4

Other data pulled for agents: 4

Multi-Agent Systems run concurrently to generate Trading Signals 5

The Commodities and Equities Agent 5

Moving Average-Crossover Agent 6

Twitter Agent 7

Decision Maker: A Posteriori Knowledge Module 7

The Decider Agent 7

Agent Interaction with the Market and the User 8

The Broker Agent 8

The PNL Agent 8

The CEO 9

Frontends 9

Zapier Frontend 9

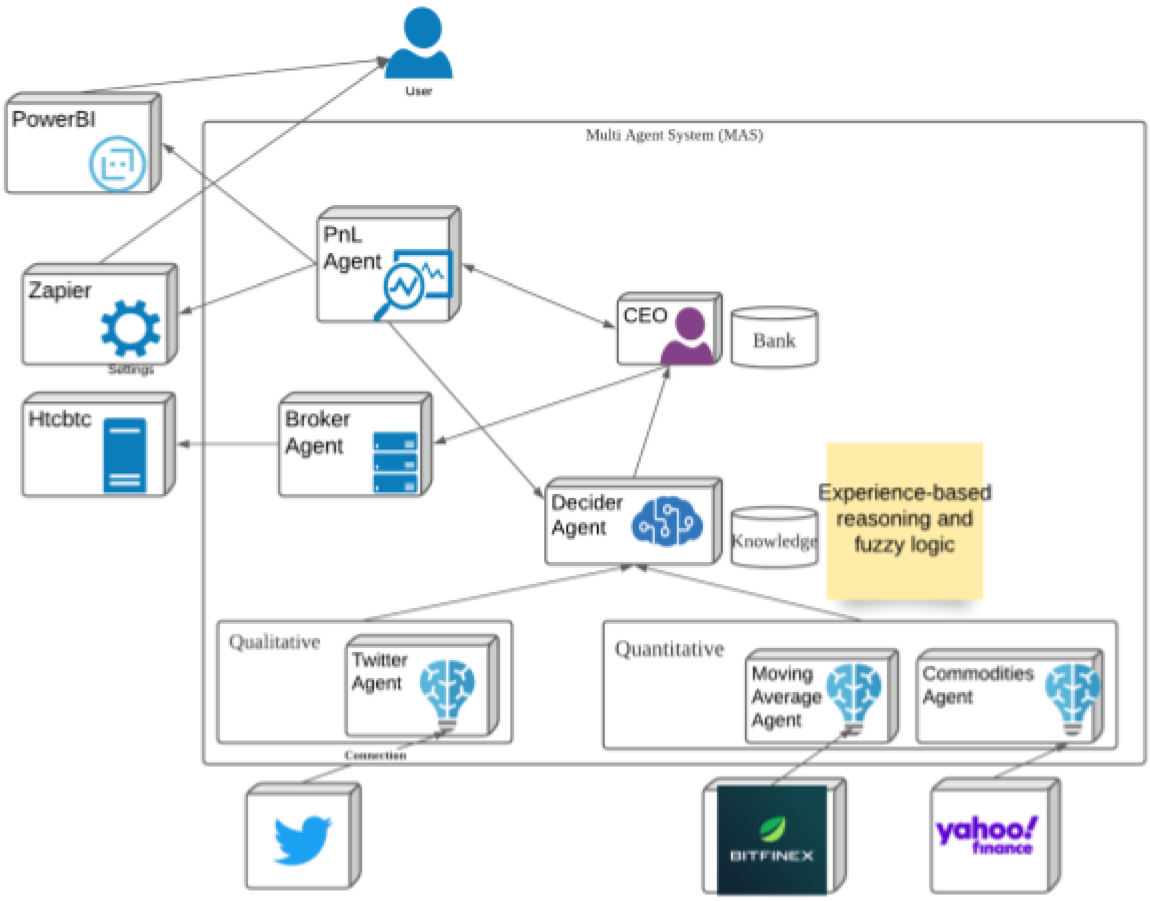
PowerBI Frontend 10

Appendix 12

This document is 1483 words excluding Cover page, Contents page, Footnotes, Tables, Captions and Appendix.

1. Overview

We developed a multi-agent system to trade EOS/USDT and ETH/USDT on HitBTC. Our agents exploit both qualitative and quantitative data to generate trading signals. A decision maker then reads these signals to make trading actions. The heart of this decision maker is a reinterpretation of CBR as a lazy learner which allows us to elegantly incorporate forward chaining and backward chaining, so as to run adaptively with changing market regimes. Finally, as several post-trade agents power our dashboard frontend, our human users receive an intuitive view of the system's live performance.



UML diagram of proposed system

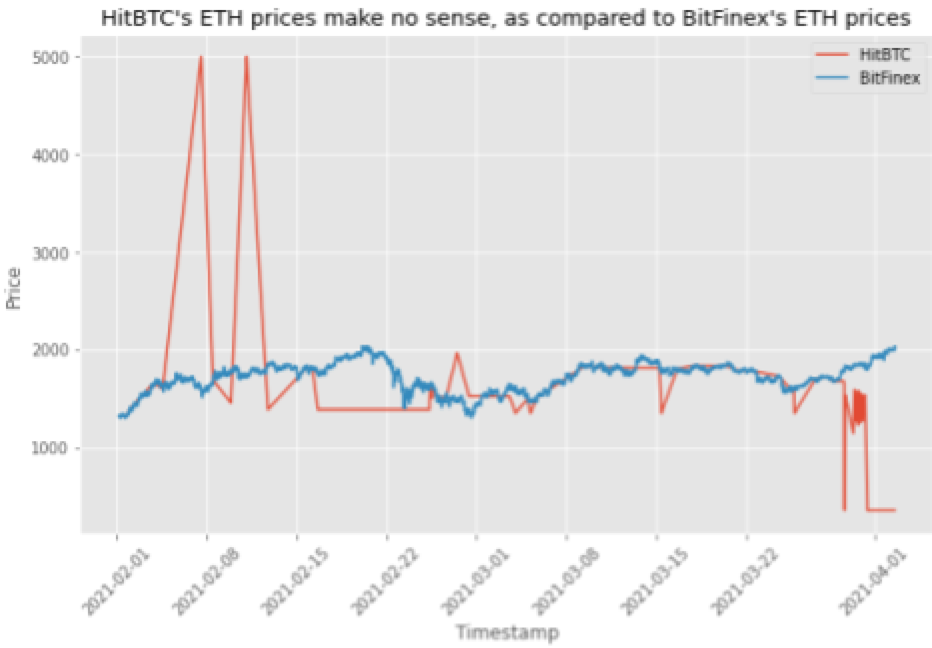
2. Our `PullingAgent` collects 3 types of data

Price data

It is a popular belief that "market action discounts everything"[1]—all predictive information is captured in price information and some scientific literature agree[2]. Thus, price data is a non-negotiable input in our system.

However, a crippling problem with price data in HitBTC's demo exchange is severe illiquidity; in ETH/USDT, only approximately 6% of all hourly timestamps are valid and its fragile orderbook is easily staggered by small trades. The graph below shows HitBTC's Etherium reaching impossible prices of 5000 as well as having very little trade activity.

On the other hand, the same ETH/USDT market in BitFinex is actively traded and there are no alarming prices there. Thus, we argue that price data retrieved from BitFinex is more informative than from HitBTC's prices and we draw our price data from BitFinex instead to train our agents.



Ethereum prices in HitBTC reach alarming prices and have very little trade activity. In contrast, BitFinex shows healthy trade activity and prices.

Other data pulled for agents:

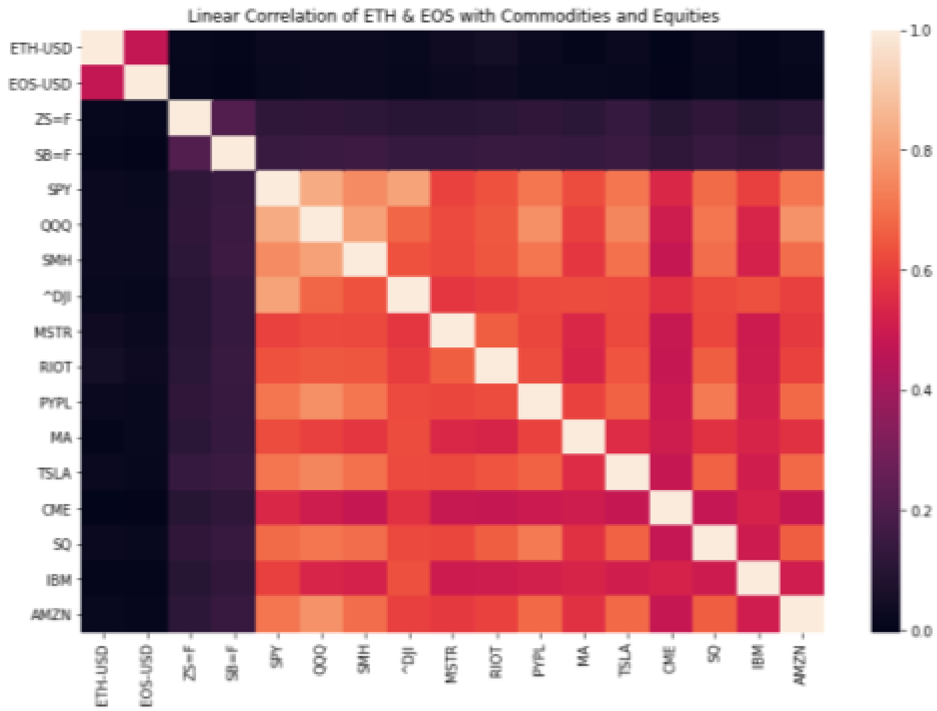
- Twitter data filtered based on hashtags
- Cross-market price data from commodity and equity markets

[1] Murphy, John J. *Technical analysis of the financial markets: A comprehensive guide to trading methods and applications*. Penguin, 1999.
[2] Neftci, Salih N. "Naive trading rules in financial markets and wiener-kolmogorov prediction theory: A study of" technical analysis"." *Journal of Business* (1991): 549-571.

3. Multi-Agent Systems run concurrently to generate Trading Signals

The Commodities and Equities Agent

Cryptocurrencies were once very attractive as financial instruments as they boast to be uncorrelated to saturated equity markets. However, the recent trends show that there is a small but increasing degree of correlation between cryptocurrencies and the commodities and equities markets[3]. We attempt to exploit price information from these markets to find profitable patterns. A list of tickers are chosen based on online articles[4], and examined with a simple correlation plot as seen below.



Correlation plot of tickers claimed to be related to cryptocurrencies. Notably, ETH and EOS remain largely, but not completely, uncorrelated with the commodities and equities markets. Close prices are used at 5 minutes interval. Ticker symbols identified in appendix.

---

[3] This article highlights remarkable but nuanced correlation between Bitcoin and S&P 500, highlighting trading strategies to trade the correlation between the two instruments, https://www.forbes.com/sites/investor/2020/05/13/bitcoin-and-stocks-correlation-reveal-a-secret/?sh=25386a9b12c2

[4] These articles highlight specific instruments that might have predictive information of future price movements in cryptocurrencies.
https://www.fool.com/investing/stock-market/market-sectors/financials/cryptocurrency-stocks/
https://www.nasdaq.com/articles/top-blockchain-stocks-to-buy-right-now-4-names-for-your-list-2021-03-30
https://www.nasdaq.com/articles/making-a-list-of-the-top-cryptocurrency-stocks-to-watch-4-names-to-know-2021-03-02
https://www.forbes.com/sites/greatspeculations/2021/02/19/with-bitcoin-at-50k-which-crypto-stocks-should-you-pick/?sh=57498adb7203

The correlation plot tells us that at a high frequency of 5 minutes interval, these instruments do not have high linear correlation with ETH or EOS. If there is any information to exploit, it must be found with a nonlinear model. Here, we follow a popular model[5]—Decision Tree Classifier.

We then frame this as a classification problem for the model. The model takes the sign of the returns of the previous days and attempts to predict the future direction of ETH and EOS. Our initial results using the daily returns were convincing.

Instrument	**ETH**		**EOS**	
Interval	**Daily**	**5 Minutes**	**Daily**	**5 Minutes**
Precision	**0.61**	0.52	**0.89**	0.51
Recall	**0.58**	0.52	**0.57**	0.50
F1	**0.56**	0.49	**0.63**	0.48
Accuracy	**0.58**	0.52	**0.57**	0.5

As apparent in the performance figures above, commodities and equities have valuable non-linear information in predicting daily price changes in ETH and EOS. *However, to ensure trades during live evaluation, we have decided to deploy the 5 minute system.*

Moving Average-Crossover Agent

We apply a moving average crossover momentum strategy on both cryptocurrencies. Specifically, these strategies use weighted moving averages which have the advantage over simple moving averages as it prioritizes the latest price information over older price data.

We further apply Chi-square tests. Future returns and the ratio of the two moving averages are discretized and binned. The binned variables are then placed through a Chi-square test. The results imply that future returns and the value of the moving averages are not independent. *In short, the ratio of the moving averages contain significant information to predict future price change.*

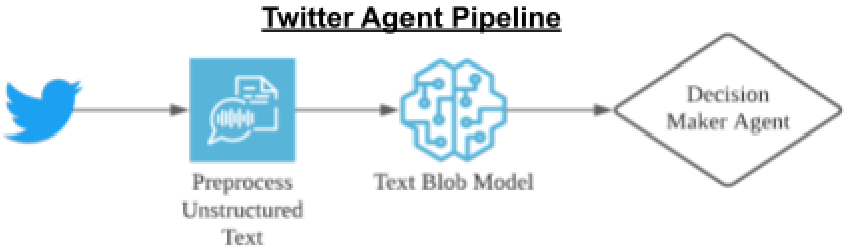
Instrument	**Chi-squared Value**	**P-value**
ETH	3.88	**0.048***
EOS	10.03	**0.000****

The signals of our moving averages are then fed into the decision maker agent.

[5] This article describes the use of decision trees to predict price changes in cryptocurrencies: https://ieeexplore.ieee.org/document/8862585

Twitter Agent

As we have learned in class that Twitter may contain predictive information that describes the sentiments of the mass over financial instruments. Our Twitter agent pulls text data from the Twitter API, applies Natural Language Processing (NLP) to generate signals to our decision maker agent.



Because social media data is noisy and highly unstructured, the pipeline needed to sensibly handle such irregularities to extract the signal from the noise. Firstly, Tweets are filtered based on hashtags such as #eosusdt, #eos, #ethusdt, #eth, #etherium and presence of financially loaded vocabulary. Secondly, the filtered tweets are further preprocessed to sharpen semantic information. The full list of string preprocessing steps and financially loaded vocabulary is found in the appendix.

The cleaned text data is fed into a textblob[6] which produces a pair of intuitive outputs—polarity and subjectivity[7]. Polarity ranges within [-1,1] where 1 means positive statement and -1 means a negative statement. Subjectivity is also a float which lies in the range of [0,1] describing if the statement is subjective and opinionative or objective and factual. The average scores of all relevant tweets in the past 30 seconds are then provided as input to the Decision Maker Agent.

4. Decision Maker: A Posteriori Knowledge Module

The Decider Agent

As can be found in our research review, Case-based reasoning (CBR) is highly related to k-Nearest Neighbours (KNN). Just like CBR, KNN is capable of retrieval, reusing, revising and retaining general knowledge for continual use as it encounters more cases. In fact, KNN is capable of all steps elegantly in a single `fit-predict` function as it is a lazy learner—by simply appending new cases to a list of historical instances and feeding it back into a single run of `fit-predict`, the decider agent elegantly initiates backward-chaining simultaneously as it performs inference, which is the culmination of forward-chaining. Thus, the KNN model constantly ranks previous cases and gives weightage to each new decision based on the profit and loss at each time step. At deployment time, the agent will start with a knowledge base stretching from 1st March to 31st March 2021 found here.

[6] The following article describes Textblobs and preprocessing steps required to handle unstructured Twitter data: https://towardsdatascience.com/twitter-sentiment-analysis-in-python-1bafebe0b566

[7] This article well describes the textblob, its uses and its outputs: https://www.analyticsvidhya.com/blog/2018/02/natural-language-processing-for-beginners-using-textblob/#:~:text=Polarity%20is%20float%20which%20lies,of%20%5B0%2C1%5D.

We frame the problem for the decider agent as a classification task. The decider agent takes in the output of the multi-agents and gives a binary predicted action of whether to open a new long position for 5 minutes or to close the long position by selling.

By coalescing the trading problem and classification task, we simplify the problem; only buy orders need to be optimized, thus *we maximize precision and recall specifically for the buy signal.* These metrics translate well to profits as a wrong sell signal results in no realized loss but a wrong buy signal does. The performance of our model can be seen below, where label 0 is the buy signal and label 1 is the sell signal. Notably, while the overall accuracy is poor primarily because of the poor performance for the sell signals, the metrics for the buy signals are very attractive.

Crypto	Label	Precision	Recall	F1	Overall Accuracy
EOS	0	0.47	0.73	0.6357	0.48
	1	0.50	0.25	0.33	
ETH	0	0.50	0.87	0.63	0.5
	1	0.5	0.14	0.21	

For optimal position sizing, we turned to the Kelly Criterion to mathematically calculate the percentage allocation of capital to maximise capital growth. Naturally, the allocation advised by the Kelly Criterion is continuously updated with the agent's knowledge. Just as the decision maker accumulates knowledge as it runs, the Decider's Agent position sizing adjusts as the model's historical performance changes.

5. Agent Interaction with the Market and the User

The Broker Agent

The Broker Agent's main job was to be an interface between all our other agents and HitBTC. It would help with executing our buy orders, square them off and gather trade and portfolio information for reporting.

In addition, we added risk management logic within the Broker Agent independent of the other agents. These were specifically to take profit and stop loss at an arbitrarily determined 20%. Since our trades happened in short intervals of 5 minutes, there wouldn't be a risk of missing out on too many momentum movements in the market.

The PNL Agent

Every bet passed from the Decider Agent will eventually be archived with the PNL agent if any orders were executed. As live environments and simulated environments may differ, the PNL agent also accounts for slippage into its calculation. It also pushes an overall model performance to the CEO for review.

The CEO

Anything can happen during trading. Upon a chance that the inconceivable happens, the human CEO role is to pull the plug on the system. As the agents constantly push figures through Zapier for human review, the CEO reserves the power to shut down the project on a whim.

6. Frontends

We give an inside look behind the trading system to users through two frontends—Zapier and PowerBI.

Zapier Frontend

Once our MAS finishes and reports are generated to a spreadsheet. We use Robotic Process Automation to eliminate any tedious and cumbersome tasks for reporting performance to the CEO. This is done by using Zapier. Whenever the MAS updates a new dataset of report into our spreadsheet, it triggers an action in Zapier. Zapier sends a html formatted email to the CEO displaying all the results and balance in the portfolio. The CEO has visibility and capability to monitor the algorithm trading's performance without the need to communicate with the software engineers.

Profit and Loss Report

Symbol	Buy Order Start	Sell Order End	Buy Trade ID	Buy Price	Buy Quantity	Sell Trade ID	Sell Price	Sell Quantity
BTCUSD	2021-04-03T16:11:43.786Z	2021-04-03T16:11:43.786Z	6826712625	53595.16	0.0001	6826712834	53229.61	0.0001

Symbol	Theoretical PnL (%)	Theoretical PNL (USD)	Actual PNL (%)	Actual PNL (USD)	Ending Balance (USD)	Rate of Return %	P.A Return %
BTCUSD	0.09570976185	0.000009570976185	-0.006820578575	-0.0000006820578575	1897.765551	-0.05111722457	-0.9999999952

Contact Us

Fintogic

We're bringing more flexibility to enhance our reccomendations for cryptocurrency trading. We welcome you to provide feedbacks on our product. Contact us at e0674589@gmail.com for more information.

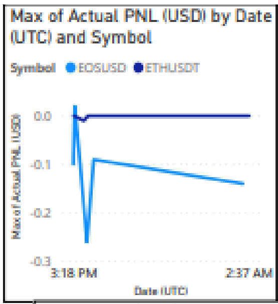
Best regards,
Group3 IS5006
Fintogic

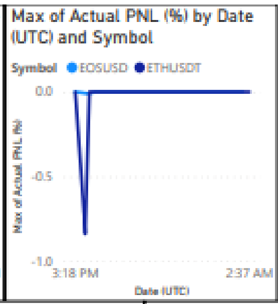
© 2021 GitHub, Inc. Terms Privacy Security Status Docs Contact GitHub Pricing API Training Blog About Loading complete

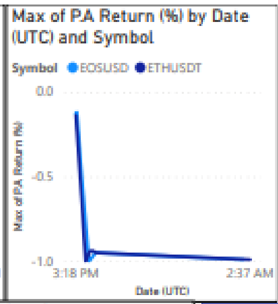
PowerBI Frontend

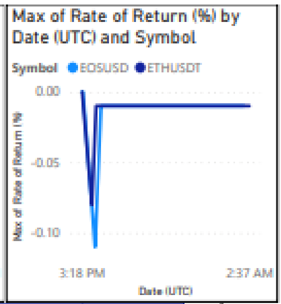
Besides the email reporting as part of the RPA process, the CEO can also request for a Data Visualization report from the Data Visualization Team. This is done with a PowerBi dashboard. The Data Visualization is a graphical representation of information and data and PowerBi provides the CEO an accessible way that is simple to see and understand the performance trend of our MAS. The MAS automatically uploads new results into PowerBI dashboard using PowerBI API and the data visualization team has live streaming of data into the dashboard. The CEO can easily access the dashboard to view live streaming of graphs and charts of the MAS performance or he can also request for a PDF view of the dashboard from the Data Visualization team.

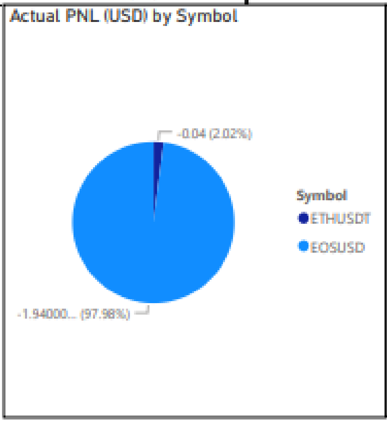
Group3 IS5006 Intelligent System Deployment Performance Of Our Multi-Agent System

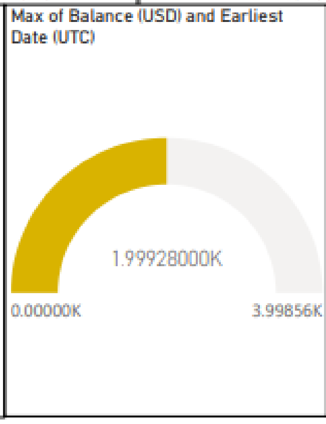












Date (UTC)	Balance (USD)
04/11/21 02:37:59 AM	1,974.28
04/11/21 02:14:08 AM	1,974.31
04/10/21 04:36:40 PM	1,983.89
04/10/21 04:36:39 PM	1,983.89
04/10/21 04:27:13 PM	1,984.70
04/10/21 04:15:10 PM	1,984.70
04/10/21 04:10:59 PM	1,775.31
04/10/21 03:58:47 PM	1,836.95
04/10/21 03:23:12 PM	1,999.28
04/10/21 03:19:56 PM	1,999.20
04/10/21 03:18:13 PM	1,999.21

Appendix

Ticker symbols used for Commodities and Equities Agent

Symbol	Instrument
ZS=F	Soybean Futures
SB=F	Sugar
SPY	S&P500
QQQ	QQQ ETF
SMH	Vaneck Semiconductor ETF
^DJI	Dow Jones Industrial
MSTR	MicroStrategy Incorporated (bitcoin orientated company)
RIOT	Riot Blockchain Inc
PYPL	PayPal
MA	Master
V	Visa
TSLA	Tesla
CME	Chicago Mercantile Exchange
SQ	Square Inc.
IBM	International Business Machines Corp.
AMZN	Amazon

Tweets are filtered by presence of financially loaded vocabulary

List of filter words for Twitter stream	
Profit	Hold / Hodl
Recommendation	trade
Lose / lost / loss	Buy
Gain	Sell
Short	price
Long	climb
Short	drop

Twitter's unstructured text is preprocessed for NLP

Steps to handle textual idiosyncrasies of Twitter data
Lowercase all string
Remove special characters, e.g. @#$
Remove hyperlinks
Remove &text, html, [video]
Remove line breaks
Remove all remaining characters that aren't letters, white space, or the following #:)(/\='] that are used in emojis or hashtags