E: Machine Learning Model: Financial Expert V1

Hi - I need a machine learning model built that interfaces to GPT to do a chat-based analysis of the dataset via Q&A. User interface will be basic for now and the data accessed via API. It will start with crypto and macro analysis datasets, and more added later. See attached writeup, and here's a short video explaining the requirements:

Please review and send me a quote for what it would take. This is the first project to be expanded upon once it's complete.

Background / Resources

- https://pub.towardsai.net/build-chatgpt-like-chatbots-with-customized-knowledge-for-your-websites-using-simple-programming-f393206c6626
- https://github.com/neefrehman/millzbot
- https://betterprogramming.pub/how-to-finetune-gpt-3-finetuning-our-virtual-mental-health-assistant-641c1f3b1ef3
- https://www.udemy.com/course/machine-learning-applied-to-stock-crypto-trading-python/?src=sac&kw=machine+learning+crypto ***(I have this course if you'd like to take it)***
- https://enhanceui.gumroad.com/l/chatgpt-starter-template ***(I also bought this if it's the best option to speed development and interfacing GPT)***
- SENTIMENT ANALYSIS: https://monkeylearn.com/blog/sentiment-analysis-of-twitter/
- SENTIMENT ANALYSIS: http://help.sentiment140.com/api

Project Description:

1. The model will be used to create a niche-expert in finance and cryptocurrency.
2. Python code will scan the internet, twitter and youtube and find the best niche content (based on likes, followers, engagement, web traffic) and use it to build a dataset. All crypto content is indexed, but low quality content is auto-excluded (very low user base, low engagement, all negative sentiment, etc). The app will have a very simple admin interface that does three things:
    a. **AUTO FLAGGING / REMOVAL**: It will green, yellow and red flag content that it thinks is of excellent, mediocre or poor quality. Poor quality is auto-removed.
    b. **REMOVAL**: The admin can then select to remove that source and it will retag it as "removed". For example, the highly regarded (with lots of positive sentiment) user on twitter https://twitter.com/WifeyAlpha would have a green flag. Some scammy bot content would be red flagged and auto removed. Anything that it can't figure out is yellow flagged.
    c. **ADD**: Allow you to add missing sources, like:
        i. a twitter profile, a certain youtube channel, a medium writer, a Reddit channel, etc.
        ii. A PDF with a book from a famous financial writer
        iii. A link to Wikipedia or a blog or a anything online
        iv. A link (plus take login info) to a course (like on udemy) where it can cycle thru content
        v. A link (plus take login info) to an ebook (like on a library website) where it can scan it
    d. **RUN**: Manually start a rerun of the dataset (if they removed or added new sources, etc).
3. Certain crypto-specific sets of data will be pulled into the dataset too via API. Here are the initial sources, but it will have an interface to add as many more as needed (*SKIP if the API is not free!*). ***NOTE: Many of these will have overlapping info (like current info on pricing, in case someone asks "what's the current price of bitcoin right now?"). Some of these might be skipped, but please let me know so i can evaluate those that are too hard to bother with right now. The idea is just to have instant access to key data available into this expert advisor's brain. But all of them have different data that someone might ask about.***
    a. SCRAPE?:

i. https://www.forexfactory.com/calendar (not an API, key data source for economic events)
ii. https://token.unlocks.app/
iii. https://github.com/Michael-M-Judd/ICO-Drops-Scraper
iv. TokenMetrics.com (login info will be provided)

b. CRYPTO PRICING/DATA:

i. https://messari.io/api (Best source for deep data on every crypto)
ii. https://docs.cryptowat.ch/rest-api/ (FREE??)
iii. https://cryptorank.io/api
iv. https://coinmarketcap.com/api/
v. https://www.coingecko.com/en/api
vi. https://nomics.com/docs/
vii. https://www.livecoinwatch.com/tools/api

c. CRYPTO QUANT DATA AND ONCHAIN

i. https://defillama.com/docs/api
ii. https://dune.com/docs/api/
iii. https://www.blockchain.com/explorer/api
iv. https://nanoly.com/api (Coindix)
v. https://docs.glassnode.com/basic-api/api
vi. https://cryptoquant.com/docs (FREE??)
vii. https://www.nansen.ai/institutions (FREE??)
viii. cbbi.info / https://www.lookintobitcoin.com/charts/ (API?)
ix. (???) https://cryptostats.community/ —> https://cryptofees.info/

d. CRYPTO SENTIMENT DATA

i. https://coinmarketcal.com/en/api (Events)
ii. https://cryptopanic.com/developers/api/ (crypto news)
iii. https://api.santiment.net/ (Sentiment) (FREE?)
iv. https://lunarcrush.com/about/api (pWeb) / https://pypi.org/project/lunarcrush/ (Sentiment)

4. The dataset will be auto- updated daily

a. Time will play a role–for example, we will be able to prompt things in the interface like “last 24 hours” or “last week” or “last month” or “last year” or “all time” for the questions being asked. The model will still use all of it’s expertise for context for all answers, but it will be able to answer about time- and date- specific requests.

5. **FUTURE**:

a. In later versions of this, the bot will be abstracted to feed in other types of expertise, so please make functions generic and separate the “financial bot and training data” from the core logic of building the model. I want to be able to create multiple expert bots by just creating a new dataset and pointing to it with a different interface (for e.g. a Healthcare bot).

b. In later versions, we will probably add other datasets (that will have their own copycat dashboard for adding/deleting sources, etc) for things like “Stock market”, “real estate”, “personal finance”, etc, to narrowly train on in addition to our first two datasets here (crypto and macroeconomics).

Example Prompts and replies:

A. **PROMPT:** “give me a 10 paragraph and bullet list summary of what happened in the last 24 hours of the crypto market, but also give good context (including the global macro) of what it means.”

a. **REPLY**: It would tell you all the major events (hacks, major coins that exploded or tanked in price and speculation as to why it happened), and maybe tell you to use caution because there is a FED meeting on Wednesday likely to cause volatility because the FED is expected to raise rates again.” Etc, etc. (NOTE: All of this and more would be available in all the analysis the bot did overnight from hundreds of tweets from the right people, hours of youtube vids that it read the transcript of, etc"

B. **PROMPT:** "Write me a newsletter about the crypto market with macro context that covers the last week in crypto from sunday Jan 1st to the 8th. Make sure it has a strong, clever intro and outro that summarizes what they will learn in the intro and what's coming in the following week in the outro."
    a. **REPLY**: Similar as the first reply but covering a whole week, written in more of a newsletter style.

C. **PROMPT:** "What is the current sentiment and situation on SOL crypto?"
    a. **REPLY**: It would give you the current general sentiment (maybe say "my sources show that SOL sentiment on Twitter, Reddit and Youtube is 68% negative") and a bullet list of what the most major SOL influencers are saying about it ("For example, even IA Crypto, who has always been a SOL bull, is not saying he's considering selling his bags of SOL if there aren't some major changes made"

D. **PROMPT:** "What are the 25 cryptos most likely to have positive price action this week, along with any caveats as to why that might not happen."
    a. **REPLY**: It would state any red flags about the week ahead–it would have this context about market risks, like the last week closed in the red, the FED meeting is coming up, BTC and the SPX are still in downtrends, etc. Then, it would pick 25 cryptos that seem to being talked about a lot, have events coming up, are in an uptrend versus BTC and the Nasdaq, etc. And it would give a reason why each one is picked (e.g. "Cardano has announced a new hard fork that everyone is talking about and that is only 2 weeks out so interest may pick up, the tweet volume about it has increased over 50% in the previous week.")

E. **PROMPT:** "Can you please write me a deep dive analysis of the QNT crypto in the following format: Tokenomics, Team, Technology, Timing " (then give descriptions of each)
    a. **REPLY**: It would give deep dives on each section, pulling from messari and all it's deep learning dataset to produce a current and timely analysis of the opportunity. <note, i'll have a specific template to train it on for this later, so it's always in the same format>

F. **PROMPT:** "What's the current price on BTC?"
    a. **REPLY**: (pulls from multiple pricing APIs to give the best answer

G. **PROMPT:** "Are there any arbitrage opportunities on any Top100 cryptos?"
    a. **REPLY**: e.g…. "Yes, there is currently a 1.2% arbitrage opportunity between the price of BNB on Coinbase and the price of BNB on uniswap."

Deliverables

1. *All code should be well documented and clean.* ***See requirements for details on all of the below.***
2. Python machine learning backend
    a. Whatever is the best database (Mongo or others) setup and running
    b. Wrapped with FastAPI to expose the results to any frontend.
    c. Assume this will be accessed vai a SaaS app so the API will require authentication.
3. Simple admin interface to review the sources that are being added to the model.
    a. This includes the red/yellow/green flagging, auto-removal of red-flags.
    b. Ability to ADD, DELETE, ReRUN model, add in new API sources, etc
4. Simple Front end for testing
    a. A very basic front end 1-page website for testing
    b. Allow all functionality to be tested (via the API)
5. Hosting (*Assume this will be an enterprise quality app requiring fast response times, etc)*
    b. Hosted in a scalable, fast, proper cloud service to be used by my app long term
    c. Should run multiple times a day to keep the data fresh