

CS425: Natural Language Communication

G1T6 - EzFinBot - Project Report

Group Members and Task Allocation

Sia Yan Rui: Query Expansion, Deployment

Yieh Yuheng: Fine Tuning, Retrieval, Reranker, Query Expansion, Evaluation

James Lim Jia Yang: EDA & Visualization, Fine Tuning, Retrieval

Lester Ye: Evaluation

Miguel: Difference with Existing Work

1. Introduction

1.1 Background

Artificial Intelligence (AI) has taken the world by storm and it will continue to do so at an exponential rate as new business models rise and new innovative technologies emerge. Those who grasp the opportunities that are made possible with AI in Finance will be the ones who stand the benefit the greatest. More than 80% of banks around the world already embrace and utilize AI in their processes, according to a Insider Intelligence AI in Banking report (Digalaki, 2021).

1.2 Goals

Institutional money has already begun harnessing the power of AI to improve their business models, but **not much of these benefits have trickled down to the everyday consumers**. We believe that on the consumers' end, there are many benefits that can be derived from generating a solution that utilizes web mining and information retrieval techniques to provide financial advice, such as **quicker response to day-to-day financial questions**, which may **aid in decision making and improve standard of living**. Hence, our group set out to create a fine-grained model using information retrieval techniques that can capture the semantic phenomena of the finance field. We will then package this model as a chatbot question and answering system that the day-to-day user can use to ask questions regarding financial decisions and get advice on.

1.3 Difference With Existing Work

The existing work utilizes a retrieval function known as BM25 to improve relevancy of search results. BM25 is used in information retrieval applications such as search engines in order to rank documents according to the occurrence of query words in each document. It was used around 2016 as a substitute for the vanilla tf-idf ranking algorithm in Wikipedia's search results (Bernhardson et al., 2016). We will be using this as a baseline approach as BM25 is a still lexical /sparse representation of text, and attempt to further improve it through dense

representations of text using embeddings. We also see similar implementations of BM25 in a retrieve and re-rank pipeline that we will be utilizing (explained more later on in methodology) that we aim to improve the implementation of and apply it to our financial QA domain (UKPLab, n.d.) .

2. Dataset

Our project utilized the FiQA-2018 dataset, a dataset focused on financial opinion mining and question answering. This dataset was built by crawling Stackexchange, Reddit and Stocktwits. There were 57368 documents, 6648 questions and 17110 mappings between questions and documents. Below is an example of question and answer pair:

Question: How to deposit a cheque issued to an associate in my business into my business account?

Answer: Have the check reissued to the proper payee.

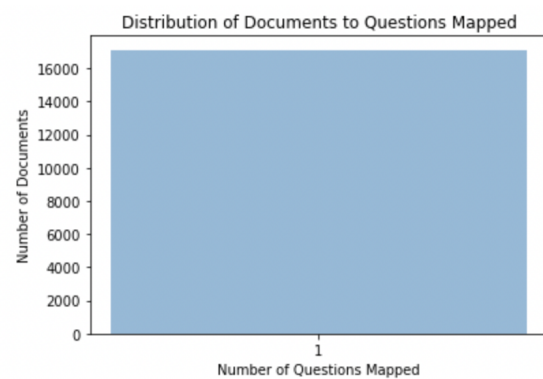
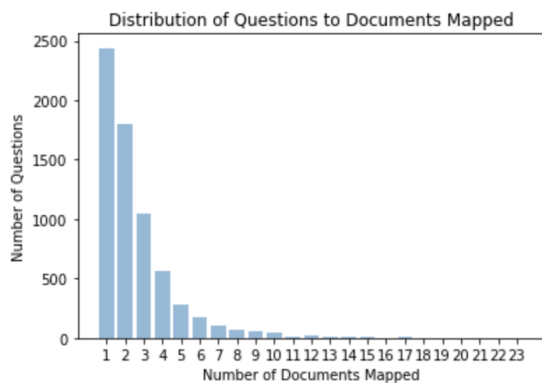


Fig 1: Distribution of questions to documents mapped Fig 2: Distribution of documents to questions mapped

Since the number of documents and questions varied by a lot, we explored the mappings between documents and questions to gain better intuition regarding their relationship. From Fig 1, we can see that there were 2439 questions that were mapped to only 1 document while the other questions were mapped to multiple documents, some up to 23 documents. Out of all the documents in the dataset, there were 40528 that were not mapped to any questions and the remaining 16480 documents were mapped to only 1 question each, as seen from Fig 2.

3. Overall Methodology

3.1 Overview

The team proposes and implements an information retrieval based question answering system. The team first performs query expansion on the user query. Next, a retrieve and re-rank pipeline is implemented. To implement this pipeline, the team evaluates a few BERT models for the retrieval system, and takes the best performing BERT model to implement a re-ranking step. The team quantitatively evaluates models based on precision@k, recall@k and F1@k. Lastly, the entire QA system is made available for use locally via CLI, and on

cloud via Telegram Bot hosted on AWS. The below diagram shows a high level overview of our methodology.

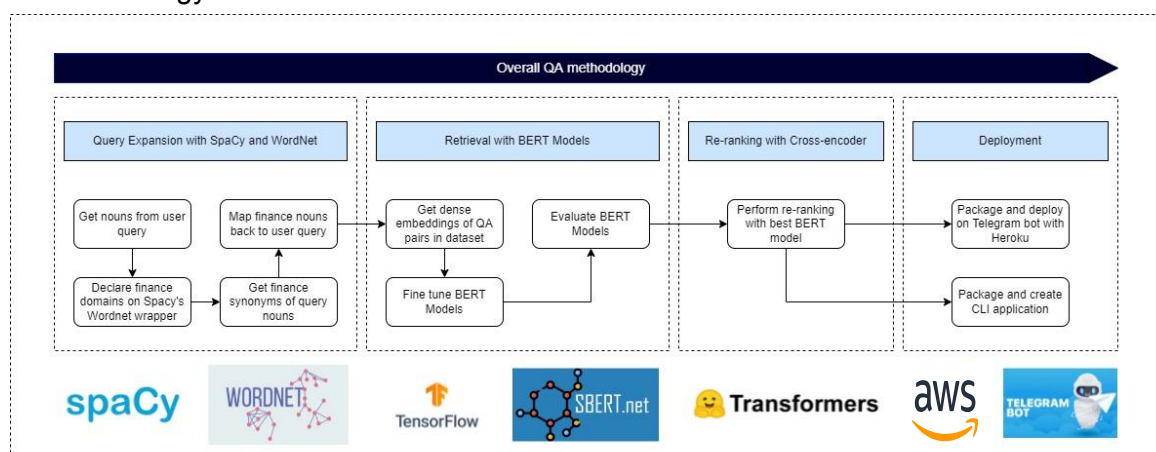


Fig 5: Overview of Methodology

3.2 Query Expansion

To improve the performance and generalizability of our information retrieval system, we will make use of ontology based query expansion in the finance domain to process the user's query. By generating synonyms of the nouns and verbs in our user's query, which generally contains most of the semantics in a user's query, we are able to improve the semantic understanding of our information retrieval system. In addition, we're able to make use of WordNet-Domains, a **lexical resource** that contains annotated synonyms within many **semantic domains**, to generate **synonyms specifically in the finance domain**. This helps to improve the accuracy of our information retrieval system.

We implemented two versions of Query Expansion. Firstly, we utilized NLTK to **identify nouns** from a given user query. We then use **NLTK Parts-of-Speech (POS) Tagging** to find the synonyms for nouns in a user query. We then map these synonyms back into the user query by replacing the original noun within a user query with "{noun} and {synonym and noun}". We only generate one synonym to prevent creating too much noise for the user's query. We also hypothesized that performance can be further improved by generating synonyms **specific to the financial domain (ontology based query expansion)**. To do so, we utilized the *spacy-wordnet* package. *spaCy* is incredibly fast compared to other text processing libraries like NLTK, as it's written from the ground up in the carefully memory-managed *Cython*, which helps us quickly process our relatively large dataset. Using the same methodology, we expand user queries with financial synonyms which will then be parsed into our biencoder model for training. An example of our query expansion:

Original: Do stock option prices predicate the underlying stock's movement?

Query Expansion Result: do stock store option monetary value prices predicate the underlying stock store s motion movement

3.3 Model Implementation

The following section details the models and pipeline utilized in the project.

3.3.1 Retrieve & Re-rank

The team utilizes a two step pipeline of retrieve and re-rank. Retrieve and re-rank is a pipeline that is well known for improving the results of information retrieval and question answering retrieval tasks.

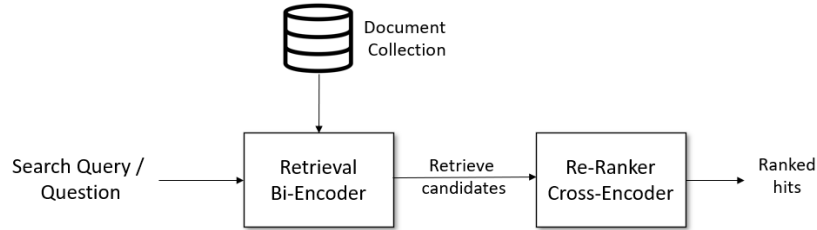


Fig 6: Retrieve and re-rank pipeline

In this pipeline, the first step is to use a **retrieval system** that retrieves a large list of possible hits which are potentially relevant for the given user query. Given this large list of possible hits, we implement a **re-ranker system** that assigns a relevance score to each potential hit with the search query. The output of this pipeline is a ranked list of hits that we can return to the user as an answer to his/her query. The rationale for such a pipeline is to **find a balance between speed and accuracy**. The retrieval system uses techniques that are less accurate than that of the re-ranker system, but is able to generate the list of possible hits faster. On the other hand, while the re-ranker system generates a more accurate score, it is unfeasible to implement it on the entire dataset. Instead, we generate a list of possible hits as a subset of the dataset, with the assumption that the ground truth exists in that subset, and use the re-ranking system to extract the ground truth. As covered in section 3.1, we first quantitatively evaluate the performance of our retrieval models to find the best performing one, before we implement the re-ranking step.

3.3.2 Retrieval Models

For the retrieval of the list of possible hits, the team utilizes a few BERT pre-trained language models to encode the user query into a dense representation (768 dimensions), and utilizes a **cosine similarity as a similarity measure** to retrieve the answers that answer embeddings are close to the encoded user query. In the case of this project, the team uses the following two BERT models: **MSMARCO-BERT-base** and **MSMARCO-distilBERT** for encoding of answers and user query, and evaluates its results qualitatively and quantitatively.

$$\text{similarity}(A,B) = \frac{A \cdot B}{\|A\| \times \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n A_i^2} \times \sqrt{\sum_{i=1}^n B_i^2}}$$

Fig 7: Cosine similarity as a similarity measure

The two BERT models here are chosen for two reasons:

Firstly, the team would like to **evaluate the accuracy between Bert-base and DistilBERT**. DistilBERT is a smaller and faster version of BERT base. By performing knowledge distillation of BERT Base, the DistilBERT model preserves over 95% of BERT's performance

on language understanding benchmark datasets such as GLUE, while being 60% faster. This can potentially **help to greatly improve user experience while maintaining quality**. Secondly, the two selected models are pre-trained already on the MSMARCO Passage Ranking Dataset. This is part of a robust method for training NLP models, called the **Transfer and Adapt** method. The idea is to **decouple learning of the task from learning of the domain to produce better results**. This fits perfectly into the project and the MSMarco dataset is a general dataset for QA/Information Retrieval while we fine-tune it further for QA/Information Retrieval in the finance domain.

3.3.3 Fine Tuning Retrieval Models

Given the retrieval models stated above, the team fine-tunes the retrieval models through a method called **Simple Contrastive Learning of Sentence Embeddings** (SimCSE). SimCSE works well for the FiQA data as this method of fine-tuning **does not require context around the QA pair to be provided**, unlike typical Information Retrieval/QA datasets such as SQuAD dataset.

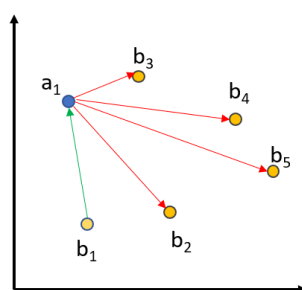


Fig 8: Representations on vector space

The concept of SimCSE is also simple to understand. Given the embeddings of all questions and answers, these questions and answers are all representations on a vector space. SimCSE optimizes the way the model embeds these questions and answers by having the correct question and answer pair (a_1, b_1) to be close in the vector space, while an incorrect question and answer pair ($a_1, b_2 \dots a_1, b_5$) should be further apart.

We quantitatively evaluate the results of our retrieval models and select the best performing one to implement re-ranking on. The best performing retrieval model is the **fine-tuned BERT-base**. In terms of retrieving a candidate set of responses, **BERT-base takes 0.105s to do so, compared to distilBERT, which takes 0.090s**. While BERT-base is slightly slower, the difference is not much, and certainly indistinguishable to a human user. Hence, the slight speed-up for using distilBERT **does not justify its drop in performance**. Hence, we will be implementing the re-ranking step with the possible hit list retrieved from our fine-tuned BERT-base model.

3.3.4 Reranker Models

Given the best performing retrieval model (fine-tuned BERT-base), we use that model to generate a possible hit list of 20 candidates, and then use the **ms-marco-miniLM-L-12-v2 cross encoder** model available on the HuggingFace library to perform reranking. The cross-encoder model takes in a user query and an answer and returns an output score between 0 to 1. The answer relevant to the user query is the one with the highest score.

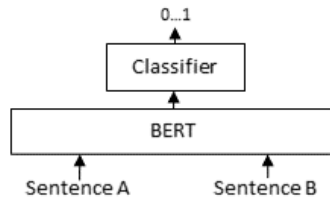


Fig 9: Cross Encoders

Cross-encoders **perform attention** across the two input sentences (in this case, the user query and the candidate hit). This results in a higher accuracy than in our retrieval step where we utilize cosine similarity on dense embeddings. However, it is also slower and hence unfeasible to implement on the entire dataset.

The team selects ms-marco-miniLM-L-12-v2 as a cross-encoder due to its high performance in the MS Marco Passage Re-ranking dataset compared to other cross-encoders (see Fig 10).

Model-Name	NDCG@10 (TREC DL 19)	MRR@10 (MS Marco Dev)	Docs / Sec
Version 2 models			
cross-encoder/ms-marco-TinyBERT-L-2-v2	69.84	32.56	9000
cross-encoder/ms-marco-MiniLM-L-2-v2	71.01	34.85	4100
cross-encoder/ms-marco-MiniLM-L-4-v2	73.04	37.70	2500
cross-encoder/ms-marco-MiniLM-L-6-v2	74.30	39.01	1800
cross-encoder/ms-marco-MiniLM-L-12-v2	74.31	39.02	960

Fig 10: Cross Encoders performance on MS Marco passage reranking

One limitation of using cross-encoder is on fine-tuning. Fine-tuning of cross-encoder is similar to training a normal machine learning model on sklearn. It requires a dataset with QA pairs, and a relevance score, which is a continuous variable between 0 and 1. However, our dataset does not provide labels, and the team is thus unable to fine-tune the cross encoder.

3.4 Deployment

As part of the team's aim to implement an end-to-end project, and for users to communicate and interact with the chatbot, the chatbot was hosted on Heroku, and deployed on Telegram. It is primarily developed with Python, using the open-source *python-telegram-bot* package, enabling us to integrate our python-based machine learning models easily. Our Telegram bot can be accessed at [@EzFinBot](#). Alternatively, the chatbot is also available for use locally via a command line interface (CLI).

3.5 Evaluation Metrics

The model results can be evaluated by calculating the Precision and Recall, before deriving the F1-Score. The F1-Score thus serves as the yardstick to compare across the different model variants, and allows us to assess whether there are significant improvements after each stage of model implementation. **Precision** gives us the percentage of Positive Cases from Total Predicted cases. **Recall** gives the percentage of how many total Positive cases

were Predicted correctly with our model. Lastly, **F1-score** gives the combined result of Precision and Recall by deriving the Harmonic Mean of Precision and Recall.

In this project, we implement a variation of these metrics. Since a chatbot only provides **one response despite there being ≥ 1 correct response**, we utilize a **metric@1 (P@1, R@1, F1@1)** approach. This approach allows us to identify if the **first entry in the result list** is present in the set of relevant answers (ground truth answers).

4. Hyper-Parameters

Fine-tuned BERTbase

attention_probs_dropout_prob: 0.1, classifier_dropout: null, hidden_dropout_prob: 0.1, initializer_range: 0.02, num_attention_heads: 12, num_hidden_layers: 12

Fine-tuned DistilBERT

Attention_dropout: 0.1, dropout: 0.1, n_heads: 12, n_layers: 6, qa_dropout: 0.1, seq_classif_dropout: 0.2

5. How to Run

5.1 Installing Requirements

The necessary requirements must be installed before running any of the files. Run *pip3 install -r requirements.txt* to install all requirements. For a more detailed guide on any possible issues you may face and how to resolve them, please refer to the README.md file.

5.2 Code Directory

The Team focused on decoupling the individual components of the chatbot. Hence, in the root directory of our project, there are 11 folders, each containing the code for the specific component of the project. The folder names describe the purpose of the folder, and running the folders **in sequence** from *1.EDA & Visualization* to *9.Deployment* will generate intermediary files required for the chatbot to work. A more detailed explanation can be viewed in the README.md files in the respective folders. These intermediary files are already generated and included in the folders for submission purposes. To skip the intermediary steps and launch the chatbot directly, please refer to *section 5.3* below.

5.3 Local CLI and Telegram Bot

To launch the terminal CLI and interact with our bot, you can run *python3 ezfinbot-cli.py* in your terminal, located in the folder "9.Deployment". To chat with our hosted bot on Telegram, you can access it via [@EzFinBot](#).

5. Evaluation Results

Model	Metrics	Result
Bm25_baseline (Using BM25)	P@1	0.264
	R@1	0.218
	F1@1	0.239
cross_encoder (Using fine tuned bertbase as retrieval, with cross-encoder for reranking)	P@1	0.404
	R@1	0.334
	F1@1	0.366
finetuned_bertbase (using fine tuned bertbase only)	P@1	0.346
	R@1	0.284
	F1@1	0.312
finetuned_distilbert (using fine tuned distilbert only)	P@1	0.332
	R@1	0.271
	F1@1	0.298
query_expansion (using fine-tuned bertbase as retrieval, cross encoders for reranking, and query expansion with NLTK)	P@1	0.387
	R@1	0.318
	F1@1	0.349
untuned_bertbase (using untuned bertbase only)	P@1	0.321
	R@1	0.261
	F1@1	0.288
untuned_distilbert (using untuned distilbert only)	P@1	0.305
	R@1	0.248
	F1@1	0.273
updated_query_expansion (using fine-tuned bertbase as retrieval, cross encoders for reranking, and domain-based query expansion with spaCy)	P@1	0.372
	R@1	0.306
	F1@1	0.336

5.1 Examples of EzFinBot's Input Queries and Responses

User> should I use a robo-advisor or a financial advisor?

EzFinBot> If your financial needs aren't complex, and mostly limited to portfolio management, consider looking into the newish thing called robo-advisers (proper term is "Automated investing service s"). The difference is that robo-advisers use software to manage portfolios on a large scale, generating big economy of scale and therefore offering a much cheaper services than personal advisor would - and unless your financial needs are extremely complex, the state of the art of scaled up portfolio management is at the point that a human advisor really doesn't give you any value-add (and - as other answers noted - human advisor can easily bring in downsides such as conflict of interest and lack of fiduciary responsibility). disclaimer: I indirectly derive my living from a company which derives a very small part of their income from a robo-adviser, therefore there's a possible small conflict of interest in my answer

User> should I invest in bonds or the stock market?

EzFinBot> Bonds provide protections against stock market crashes, diversity and returns as the other posters have said but the primary reason to invest in bonds is to receive relatively guaranteed income. By that I mean you receive regular payments as long as the debtor doesn't go bankrupt and stop paying. Even when this happens, bondholders are the first in line to get paid from the sale of the business's assets. This also makes them less risky. Stocks don't guarantee income and shareholders are last in line to get paid. When a stock goes to zero, you lose everything, where as a bondholder will get some face value redemption to the notes issue price and still keep all the previous income payments. In addition, you can use your bond income to buy more shares of stock and increase your gains there.

User> what platform should I use to trade on the stock market?

EzFinBot> You can have a look at betabrokers. It's an simulated stock trading platform which is entirely email-based. You start with 10 000\$ and you make transactions with commands in the subject line of the email (e.g. "buy 250\$ AAPL" or "cover 20 shares of AAPL"). It should be straightforward to add an email interface to your python script.

User> I'm 24 years old now, should I start investing now?

EzFinBot> As AskAboutGadgets notes, there's no lower age limit. You current age (24) is a pretty good one; you'll have four decades or so for your money to grow and compound, allowing it to become a veritable fortune when you're ready to retire if you invest it fairly aggressively.

User> should I buy a new car or a new house?

EzFinBot> Buy a house when you can, but keep driving your current car until it dies. In ten years' time, a house should be worth more than you paid for it, while a new car will be worth next to nothing. And research shows that buying possessions like cars doesn't actually make you happier, even though you think it will.

User> how can I reduce the amount of taxes that I'm paying

EzFinBot> Consider the individual who pays \$1,000,000 in taxes. His/her income must be substantial. That is what one should aim for. Investments for the most part, do not lower ones taxes. In one of John Grisham's novels, tax shelters are being discussed. Sorry, I do not remember which book. The discussion goes something like this: There are a few investments which can lower your taxes. Purchase a house. Mortgage interest on your principle residence is deductible (if you itemize deductions). If you don't itemize, focus on increasing income to the point where itemizing benefits you. In general, businesses have more deductions than individuals. Own a small business. You (or your accountant) will discover many deductions. Hint: the company should lease a car/truck, many meals are now deductible. This is not the reason to own a business.

User> what other avenues of investments can I invest in

EzFinBot> There's already an excellent answer here from @BenMiller, but I wanted to expand a bit on Types of Investments with some additional actionable information. You can invest in stocks, bonds, mutual funds (which are simply collections of stocks and bonds), bank accounts, precious metals, and many other things. Discussing all of these investments in one answer is too broad, but my recommendation is this: If you are investing for retirement, you should be investing in the stock market. However, picking individual stocks is too risky; you need to be diversified in a lot of stocks. Stock mutual funds are a great way to invest in the stock market. So how does one go about actually investing in the stock market in a diversified way? What if you also want to diversify a bit into bonds? Fortunately, in the last several years, several products have come about that do just these things, and are targeted towards newer investors. These are often labeled "robo-advisors". Most even allow you to adjust your allocation according to your risk preferences. Here's a list of the ones I know about: While these products all purport to achieve similar goals of giving you an easy way to obtain a diversified portfolio according to your risk, they differ in the buckets of stocks and funds they put your money into; the careful investor would be wise to compare which specific ETFs they use (e.g. looking at their expense ratios, capitalization, and spreads).

User> should I buy a HDB resale or a condominium resale unit?

EzFinBot> If you buy a townhouse, you often are in a condominium arrangement in the US (when you're really in a rowhouse in particular). So that's a downside right away: you have to have a HOA, or at least some sort of common agreement, though it might not have formal meetings. Everyone who owns an interest in the entire group of townhouses gets some say in landscaping and such. Beyond that, though, townhouses (and similarly, condominiums) are often easier to own (as they don't have as much maintenance that you have to do), but more expensive because you pay someone to do it (the landscaping, the external repairs, etc.). You likely don't have as much control over what the external looks like (because you have to be in agreement with the other owners), but you also don't have to do the work, unless your agreement is to collectively do the mowing/landscaping, which you should know in advance. I wouldn't underestimate the value of easier, by the way: it's very valuable to not have to deal with as many repairs and to be able to go a week without thinking about mowing or watering. In that sense it can be a nice transition into ownership, getting some-but-not-all of the obligations. But if that's something you really value, doing the landscaping and mowing and whatnot, that's relevant too. You can always tell your realtor to look for townhouses where the owners do some/all of the landscaping, though that opens up a different can of worms (where you rely on others to do work that they may not do, or do well). They're also somewhat noisier; you may be sharing a wall (but not necessarily, air-gap townhouses do exist) and either way will be closer to your neighbors. Does noise bother you? Conversely, are you noisy? In a college town this is probably something to pay attention to. Price wise, of course stay well within your means; if being close to the city center is important, that may lead you to buy a townhouse in that area. If being further out isn't a problem, you'll probably have similar choices in terms of price as long as you look in cheaper areas for single family homes.

User> is microsoft a good stock to invest in?

EzFinBot> Microsoft wants to buy a majority in the stock. To accomplish that, they have to offer a good price, so the current share owners are willing to sell. Just because the CEO of LinkedIn agreed to the deal doesn't really mean much, only that he is willing to sell his shares at that price. If he does not own 50%, he basically cannot complete the deal; other willing sellers are needed. If Microsoft could buy 50% of the shares for the current market price, they would have just done that, without any negotiations. That is called a hostile take-over.

User> what are some blue chip stocks that is a good investment?

EzFinBot> Well Company is a small assets company for example it has 450,000,000 shares outstanding and is currently traded at .002. Almost never has a bid price. Compare it to PI a relative company with \$50 million market cap brokers will buy your shares. This is why blue chip stock is so much better than small company because it is much more safer. You can in theory make millions with start up / small companies. You would you rather make stable medium risk investment than extremely high risk with high reward investment I only invest in medium risk mutual funds and with recent rollout I made 182,973 already in half year period.

User> is insurance something essential to have?

EzFinBot> Most people buy insurance because it is legally required to own a car or to have a mortgage. People want to own homes and to have personal transportation enough that they are willing to pay for required insurance costs. There are a lot of great explanations here as to why insurance is important and I don't want to detract from those at all. However, if we're being honest, most people are not sophisticated enough to measure and hedge their various financial risks. They just want to own an home and to drive a car.

User> how many percent of my income should I save?

EzFinBot> A person who always saves and appropriately invests 20% of their income can expect to have a secure retirement. If you start early enough, you don't need anything close to 20%. Now, there are many good reasons to save for things other than just retirement, of course. You say that you can save 80% of your income, and you expect most people could save at least 50% without problems. That's just unrealistic for most people. Taxes, rent (or mortgage payments), utilities, food, and other such mandatory expenses take far more than 50% of your income. Most people simply don't have the ability to save (or invest) 50% of their income. Or even 25% of their income.

User> how much do I need for retirement in singapore?

EzFinBot> I'll offer another answer, using different figures. Let's assume 6% is the rate of return you can expect. You are age 25, and plan to retire at age 65. If you have \$0 and want \$1M at retirement, you will need to put away \$524.28/month, or \$6,290.40/year, which is 15% of \$41,936. So \$41,936 is what you'd need to make per year in order to get to your target. You can calculate your own figures with a financial calculator: 480 months as your term (or, adjust this to your time horizon in months), .486755% as your interest (or, take your assumed interest rate + 1 to the 1/12th power and subtract 1 to convert to a monthly interest rate), 0 as your PV, and \$1M as your FV; then solve for PMT.

User> is the stock market bubble going to burst soon?

EzFinBot> This refers to the faulty idea that the stock market will behave differently than it has in the past. For example, in the late 1990s, internet stocks rose to ridiculous heights in price, to be followed soon after with the Dot-Com Bubble crash. In the future, it's likely that there will be another such bubble with another hot stock - we just don't know what kind. Saying that "this time it will be different" could mean that you expect this bubble not to burst when, historically, that is never the case.

User> how much return of investment should I earn to beat annual inflation?

EzFinBot> You should invest your money. To figure out what rate of return you need, use this equation: (How Much Money You Want Per Year) / (Total Amount of Cash You Have) = (Annualized Interest Rate) If we plug in the amount of annualized interest you can expect to safely get while not managing your money personally, 2% by my estimate, we get $X / 1.2m = 0.02\%$; $X=24k/\text{year}$ A measly \$24,000 / year. Many people say that you can get 10-12% even 30% return on your investment. I won't speculate on if this is true, but I will guarantee that you cannot get those returns simply by handing your money over to a money manager. So your options are, 1) Earn a guaranteed \$24,000 and earn the rest you need to live by working 2) Learn to invest your money (and then do so intelligently) and earn enough to live off the interest To learn how to invest your money, read Beating the Street, by Peter Lynch. <https://www.amazon.ca/Beating-Street-Peter-Lynch/dp/0671891634> Good luck!

5.2 Analysis of Evaluation Results

From the **quantitative** results above, we can observe a few points on the outcome. Firstly, compared to BM25, **dense representations of text works better** for our QA IR system, which indicates that the meaning of words is important, rather than just looking at words lexically. **Fine-tuned BERT models and the implementation of cross-encoders also improves performance.** On the other hand, both implementations of query expansion resulted in a disimprovement of performance. This could indicate that the decrease in performance may be **due to the noise added in expanding nouns.** An example of this could be because of a **poorer sentence structure after expansion.** This goes against the previous hypothesis that query expansion increases performance because of generating synonyms that aren't within the financial domain. Overall, while quantitative results may not look very high, **the chatbot works decently from a qualitative perspective,** and this may be due to **high semantic similarity between candidate answers.**

From the **qualitative results,** we see that the bot performs quite well. We also see in the 14th example that the bot is **able to tie the idea of “stock market bubble” to soaring stock prices and give an example** of the Dot-Com Bubble crash. However, the 15th example is an example where the bot **failed to detect the question’s focus on annual inflation.** In the 13th example, we see that the bot can **understand the intent of calculating money for retirement,** but **fails to tailor for a Singaporean context.** This can be because the data was scraped from social media sites such as Reddit, which see **a large non-Singaporean user base,** resulting in answers **more tailored for other countries.**

5.3 Qualitative Comparison with Existing Enterprise QA Bot (SAP Conversational AI)

We also deployed another telegram chatbot [@SAPCAIFinBot](#) trained on the same FiQA dataset to evaluate our bot against it. Our chatbot performed comparatively better in **identifying relevant answers,** as shown in our presentation. However, the SAP bot performed better in **making conversation,** as it recognises greetings and small talk.

6. References

Digalaki, E (2021, Jan 14). The impact of artificial intelligence in the banking sector & how AI is being used in 2021. Business Insider.

<https://www.businessinsider.com/ai-in-banking-report>

Amati G. (2009) BM25. In: LIU L., ÖZSU M.T. (eds) Encyclopedia of Database Systems. Springer, Boston, MA.

https://doi.org/10.1007/978-0-387-39940-9_921

Bernhardson, E., Causse, D., Jones, T., Popov, M., Tankersley, D., & Xie, C. (2016, October 11). Improving search result relevancy on Wikipedia with BM25 ranking.

<https://wikimedia-research.github.io/Discovery-Search-Test-BM25/>

UKPLab. (n.d.). *Retrieve_rerank_simple_wikipedia*. Google Colab.

https://colab.research.google.com/github/UKPLab/sentence-transformers/blob/master/examples/applications/retrieve_rerank/retrieve_rerank_simple_wikipedia.ipynb