FINNY - A STUDENT LOAN CHATBOT

Ureem James

MSDS 453, Fall 2023, Section 57

Northwestern University, Natural Language Processing

November 30th, 2023

**Introduction**

As of 2023, Americans owe over approximately $1.8 trillion in Student Loan Debt. Nick Ducoff, Founder and CEO of Edit, argues that the primary contributing factor has been the lack “...of adequate education, information or support systems in place to help students (and their families) make smarter college choices while they are still in high school” (Ducoff, 2019).

That said, even though the lack of information around college financing and tuition costs hurts all students, it often hurts first-generation college students even more given that they lack the social-cultural capital in their surroundings to make informed decisions. Finny, a chatbot, aims to resolve this issue by making information regarding federal student loans and financial aid more accessible.

**Data Sources**

Eight various wikipedia pages were used to develop Finny’s knowledge base. Furthermore, when processing the text into one corpus, the page sections highlighted below were removed as they were producing considerable noise in preventing the models arrive at the correct answers within the corpus:

* References
* See Also
* Further Reading
* External Links

**Question Design and Modeling Methods**

Performance Metrics

Before we dive into the various modeling techniques tested whilst developing Finny, it’s important to first frame how we want Finny to interact with its users as that will be crucial in concluding upon the final model of choice.

Given that Finny will primarily be interacting with students, its efficacy will be diminished if it’s just an information retrieval system. Humans have the strong ability to ask the same question with different word choices. A question like “*What are the consequences of defaulting on my loans?”* can also be rephrased informally as *“What if I can’t pay my loans back?”*. A suitable model for Finny will be one that can equate the two and respond from the text accordingly.

Methods

It comes as no surprise then that a suitable model must have rich embeddings to be able to conclude that the two questions semantically are asking the same question. Therefore, models that are heavily reliant on the choice of words in the question being paired up with the choice of words in the corpus, might not be suitable for our purpose. Having said that, we will test the four model techniques outlined below (some of which are dependent on keyword querying):

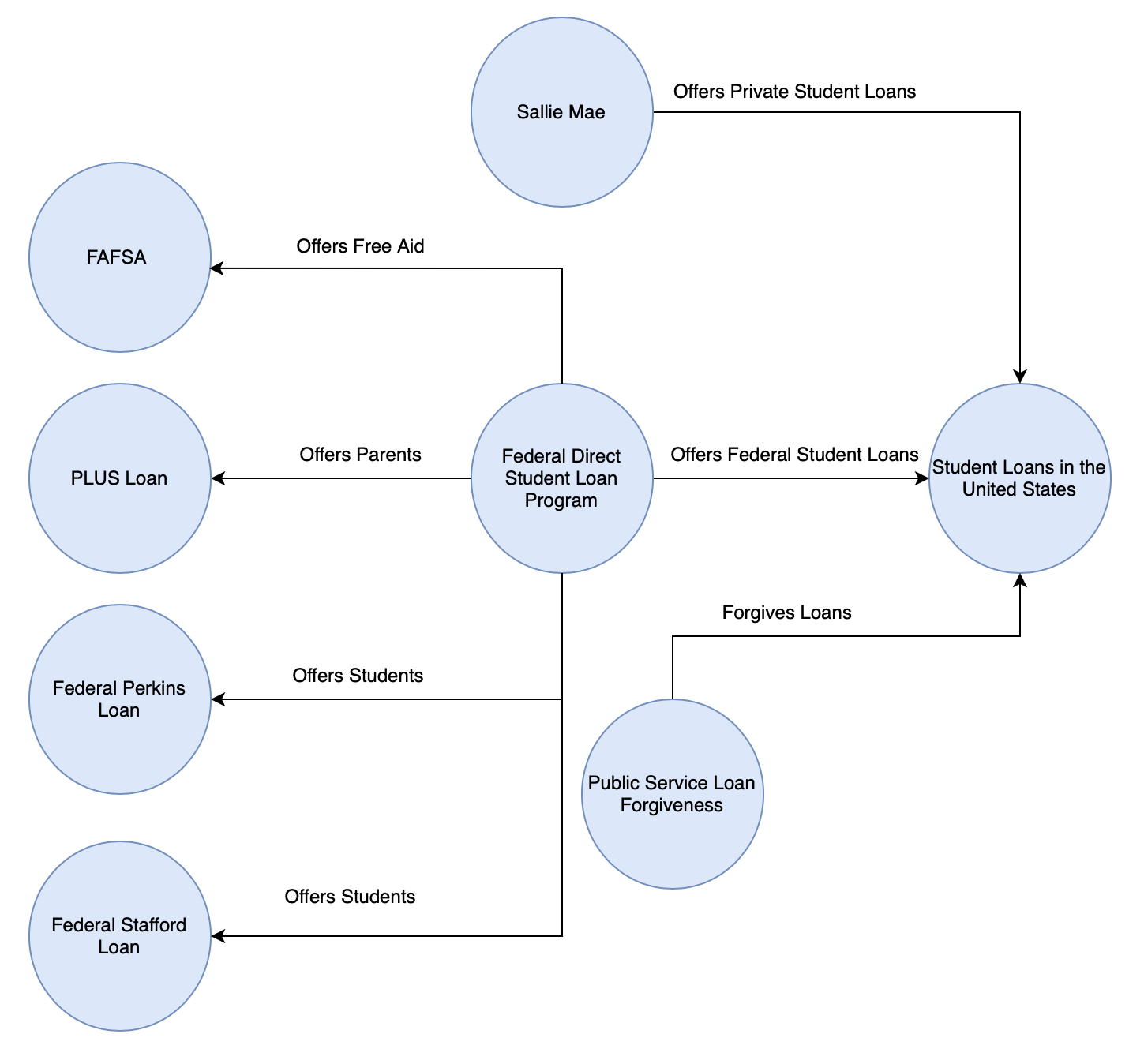
* TF-IDF
* Sentence Transformers
* DistilBERT
* Open AI GPT 3.5 (Llama Index)

Question Design

As previously stated, our final selected model will need to just be accurate in the information it retrieves, but must do so by being semantically astute in equating questions that are formally posed and informally posed as well. Consequently, we have designed a list of six questions that accomplish this goal:

* What is FAFSA?
* Am I eligible to apply for FAFSA?
* What is the difference between subsidized and unsubsidized loans?
* What is the maximum amount of money I can receive in Pell grants?
* What if I can’t pay back my loans?
* Are there any loan forgiveness programs I can apply for?

**Ontology**

Before we proceed to apply various modeling techniques onto our corpus, it might be helpful to see if we can visualize an ontology based on our various documents to understand the relationship between our documents -

**Results & Analysis**

Overall, the Open AI GPT 3.5 model provided the most accurate results whilst also being able to respond to questions both formally and informally posed. Kindly refer to the section **Criticism of** **Winning Model Responses** to review GPT 3.5 model’s responses to all the questions. Given that there are four model techniques being tested on six different questions, instead of showcasing each and every model’s response to each question, I narrowed down the results to elaborate on the performance of each model on one essential question - *“What if I can’t pay back my loans?”*

This is quite an integral question to the student loan debt debacle given that most borrowers have found it incredibly difficult in paying back their loans. The purpose of this question is to make the borrower aware of the consequences of student loan default.

The GPT model’s response to the query was quite thorough as it outlined each and every consequence highlighted in the corpus:

If you are unable to pay back your loans, there are several potential consequences. These can include damaged credit, ineligibility for future student loans, garnishment of wages, high collection fees, loss of federal income tax refunds or Social Security, and prohibition from other federal assistance programs. It is important to note that defaulting on student loans can have severe financial and legal implications. It is advisable to explore options such as income-driven repayment plans or seeking assistance from loan servicers to avoid defaulting on your loans.

Furthermore, the **GPT model** was also able to pair the consequences of default with additional information like income-driven repayment plans which reconfigure the loan payments to the borrower’s income and increase in proportion to the borrower’s income.

Additionally, **two Sentence Transformer models** were also tested - 1) a language model tuned for **semantic search** and 2) a smaller **all purpose language model**. Both models locate their answer using cosine similarity and choose the second highest ranking response from sentences available; the highest ranking being the question itself. Altogether, with regards to the query, both models performed similarly. The semantic search model responded by outputting - “if the student returns to least half-time status, the loans are again deferred, but a second episode no longer qualifies and repayment must begin”. Whereas the all purpose model responded by outputting - “loan amounts are limited”.

Though both the models were unable to answer the question, the difference between the two responses is quite stark. For the most part, at the very minimum the model tuned for semantic search was at least able to conjure some relatively coherent response, whereas the all purpose model wasn’t able to do so. This could very well be due to how each model is processing the sentences and leveraging its pre-existing embeddings; it’s very likely that the smaller all purpose model, given that it was probably trained using fewer parameters, isn’t as rich with its embeddings in comparison to the model tuned for semantic research.

That said, if we reverted the question back to “*What are the consequences of defaulting on my loans?”*, both the models, even the smaller all purpose model, responded by outputting - “as previously mentioned, default consequences are severe and can include damaged credit, ineligibility for future student loans, garnishment of wages, high collection fees, loss of federal income tax refunds or social security and prohibition from other federal assistance programs.” This response highlights the dependency of both models on certain keywords being part of the question for the model to provide accurate responses - a trait not suitable for our purpose.

The DistilBERT model is a distilled version of the BERT model and has 40% less parameters than the original model providing an increased speed in processing. That said, the DistilBERT model with regards to our query performed fairly poorly by providing the response - “the cost of going to school is not worth the debt”. If we reverted back to the formal line of questioning, the model responded with the following output - “and the consequences are much more severe”. Even though both responses didn’t meet the mark, we can see with the second response that the model was at least able to narrow down to a much more appropriate response. The response was deemed appropriate because the sentence in the corpus, that each model is trying to access in their own way, starts with “As previously mentioned the consequences of default are severe and can include…”.

We’re somewhat experiencing the same phenomenon we did with the sentence transformers - given that these models are smaller, they aren’t able to enrich their embeddings using our corpus far enough to be able to semantically understand informal queries. Instead, they’re relying heavily on keyword matching to the degree that they can only understand the formal query, and even then are stopping after equating **default consequences** to **severe** instead of listing the consequences.

The TF-IDF technique is perhaps the modeling technique that overtly utilizes keyword matching in its attempt to generate responses. When posed with the informal version of the query, the TF-IDF model responded by outputting - “once aggregate limits are met, the student is ineligible for additional stafford loans until they pay back a portion of the borrowed funds”. Whereas, when presented with the formal iteration of the question, the model responded with a one word answer - “severe”. This response makes sense as the TF-IDF model too likely equated the terms **default consequences** and **severe** and left it at that.

**Criticism of Winning Model Responses**

Below are the Open AI GPT model’s responses to the remaining five questions:

1. *“What is FAFSA?”*

FAFSA stands for the Free Application for Federal Student Aid. It is a form completed by current and prospective college students in the United States to determine their eligibility for student financial aid. The FAFSA collects demographic and financial information from students and their parents/guardians, and the results of the application are used to determine the student's eligibility for various types of federal financial aid programs, such as grants and loans.

1. *“Am I eligible to apply for FAFSA?”*

To determine your eligibility for FAFSA, you must meet certain criteria. These criteria include maintaining a satisfactory academic progress, being a U.S. citizen, U.S. national, or eligible non-citizen, having a valid Social Security number, having a high school diploma or GED, signing the certification statement regarding federal student loans and grants, not being found guilty of the sale or possession of illegal drugs while receiving federal aid, and, for male students between 18 and 25, being registered with the Selective Service System. It is important to note that eligibility requirements may change, so it is recommended to check the official FAFSA website or contact the U.S. Department of Education for the most up-to-date information.

1. *“What is the difference between subsidized and unsubsidized loans?”*

Subsidized loans are a type of federal student loan where the government pays the interest while the student is enrolled in school at least half-time, during the grace period, and during deferment periods. This means that the loan balance does not increase during these periods.On the other hand, unsubsidized loans are also federal student loans, but the interest accrues and is added to the loan balance while the student is in school and during other periods. This means that the loan balance increases over time with unsubsidized loans.

1. *“What is the maximum amount of money I can receive in Pell Grants?”*

You can determine if you are eligible to receive a Pell Grant and the maximum amount you can receive by completing the Free Application for Federal Student Aid (FAFSA). The FAFSA is used to assess your financial need and determine your eligibility for various types of financial aid, including the Pell Grant. The maximum amount of the Pell Grant can vary each year and is determined by factors such as your Expected Family Contribution (EFC) and the cost of attendance at your chosen college or university. It is important to complete the FAFSA accurately and on time to ensure you are considered for the maximum amount of financial aid for which you may be eligible.

1. *“Are there any loan forgiveness programs I can apply for?”*

Yes, there are loan forgiveness programs that you can apply for. One such program is the Public Service Loan Forgiveness (PSLF) program. Under this program, individuals who work full-time for a qualifying employer and make 120 on-time, monthly payments under a qualifying repayment plan may be eligible to have the remaining balance of their eligible student loans forgiven. It is important to note that there have been some issues and challenges with the PSLF program, as mentioned in the context information. Therefore, it is recommended to thoroughly research and understand the eligibility criteria and requirements of any loan forgiveness program before applying.

Even though it seems that the model responded quite well to all the responses, for questions four and five, the model provided limited responses. Within the text, the reader can actually locate the maximum Pell grant that was awarded for the year 2019-2020, which was around $6000. Granted, that information wouldn’t be applicable to current users, but it would be helpful for users to know the ballpark amount that the maximum Pell grants hover around.

Additionally, in the final response regarding loan forgiveness programs, the model failed to capture any reference to the teacher's loan service forgiveness program mentioned within the text given that information regarding the teacher’s loan service forgiveness program would be very useful to students looking to pursue teaching careers. It’s likely that either the text would be need to be further cleansed, or the model would have to be trained with additional prompts to extract the information required.

**Conclusions and Further Steps**

In this paper, we set out to develop a chatbot to assist students and highschoolers accessing information related to federal student loans and financial aid. That said, we also emphasized that a successful chatbot must not only be an accurate information retrieval system, but also a system that is able to process informal usage of language as the primary users of this chatbot will be highschool students.

With that specific purpose in mind, the Open AI GPT 3.5 model performed the best as it was not only able to provide accurate information, but also able to demonstrate a rich semantic understanding of informal and broad usage of language. The other models, on the other hand, though were able to arrive at reduced answers using formal language, failed to process informal usage of language.

That said, though the model contains enough information to assist students in learning more about different kinds of federal loans, financial aid, Sallie Mae and default rates, it does not contain any information regarding potential scholarships that students could apply for or information related to college costs. I’d like to consider including these additional pieces of information as I think it could provide additional value to not just students, but also their household members that are assisting them in their journey to financing a college degree.

**Appendix A**

* Student Loans in the United States
* Student loan default in the United States
* Public Service Loan Forgiveness
* FAFSA
* Federal Stafford Loan
* Federal Perkins Loan
* PLUS Loan
* Sallie Mae

References

Fry, Richard. “First-Generation College Graduates Lag behind Their Peers on Key Economic Outcomes.” Pew Research Center’s Social & Demographic Trends Project, May 18, 2021.https://www.pewresearch.org/social-trends/2021/05/18/first-generation-college-graduates-lag-behind-their-peers-on-key-economic-outcomes/.

Nick Ducoff, CEO and founder of Edmit. “Students and Their Families Need Support Systems before They’re Saddled with College Debt.” CNBC, May 17, 2019. https://www.cnbc.com/2019/05/16/students-families-need-support-systems-to-handle-college-debt-crisis.html.

**Conclusions and Future Work**