Personalizing the Digital Experience: Using Machine Learning Models for an Appealing User Experience

Politics of Designing a Digitally Personalized Experience

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia, Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By

Kaleb Getachew

Spring 2022

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

Advisor

Sean Ferguson, Department of Engineering and Society

Personalizing the Digital Experience: Using Machine Learning Models for an Appealing User Experience

CS 4991 Capstone Report, 2022

Kaleb Getachew
Computer Science
The University of Virginia
School of Engineering and Applied Science
Charlottesville, Virginia USA
kg9rv@virginia.edu

Abstract

PayPal, a financial technology company, operating as one of the premiere online payment systems, desired to increase its overall customer satisfaction of quality from 77%. By comparison, many of PayPal's competitors had satisfaction scores well above 85%. To match or exceed these scores, PayPal needed to increase their quality within their products. One effective solution is to personalize their website in such a way that users feel accommodated and welcomed to use the product. I was assigned to the World Ready Team, whose main goal was to ensure the product was well prepared for worldwide use. The team leveraged the latest models and tools within the world of Machine Learning (ML) and Natural Language Processing (NLP), to build the infrastructure and language models for the creation of a recommender systems. We first created a high-level pipeline of the entire project. We utilized a company server for storage and PostgreSQL for database management. Most of the work relied heavily on the industry standard for ML and Data Science, JupyterNotebooks. In addition, we also used several frameworks, including NodeJS and React, to establish and disseminate our work at the end of the project. Future work includes fine-tuning models, and gathering diverse corpora to better train the models to handle robust data. In addition, altering the API to become more scalable would allow for a much easier experience when dealing with the tools of this project.

1. Introduction

Every day, people are faced with decision-making situations countless times, many of which are unnoticed. Naturally, experts or friends are the first point of contact to aid in decision-making, but with the widespread use of the internet within the 21st century, the transfer of information has become very rapid, leading to the tracking and storing of user data for data-driven personalization.

Throughout the summer of 2021, my responsibilities as a Software Engineering Intern within the World Ready Team at PayPal included all things related to preparing the product for the world. I worked with two other interns and various other professionals from different departments to ensure users from all sectors were tended to and had a meaningful experience while interacting with the company's suite. The team worked specifically on converting, preparing, and personalizing Polish data for live use this year. The overarching task was to improve user satisfaction whenever users would visit the page, through the use of relevant NLP models. This meant ensuring everything displayed is easily understandable, accessible, and most of all a relatable format for each user. All of this was achieved based on data collected on each individual user.

2. Background

At the core of creating a personalized digital experience, recommender systems play a huge role in engaging users. With the recent advances in AI, data analytics and big data over the past years, opportunities have arisen for recommender systems to embrace the impressive achievements of AI. Recommender systems create advanced insights into the relationships between users and items, presenting more complex data representations, and discovering comprehensive knowledge in demographical, textural, virtual and contextual data. Big Tech Companies (BTCs) like Netflix create a personalized web/digital-experience based on consumer collected data to intentionally improve overall customer satisfaction among other unintended consequences. The Oxford Academic definition of personalization is the “strategic creation, modification, and adaptation of content and distribution to optimize the fit with personal characteristics, interests, preferences, communication styles and behaviors” (Bol, 2018). This paper will investigate the effect of data collection and data usage by BTCs in the creation of an imaginary personalized digital experience.

3. Related Works

Bol’s 2018 paper revolving around personalization of websites and platforms discusses the potential effects of self-disclosure across health, news and commerce outlets. She provides numerical proof, taken from a study, which suggest that people view the personalization effect in different light depending on the industry at hand. “Diversity by Design in Music Recommender Systems” by Porcaro, Castillo, and Gomez (2021), explores the effect and power of recommender systems within the context of the Music industry. They state that humans learn through various sources of information, and that these recommender systems have a responsibility to diversify the content of music they recommend to their users. This relates very well to the film/streaming industry explored within this paper. The article titled “Selecting the best artwork for videos through A/B testing” a Netflix Technology Blog discusses, discusses the technicalities that Netflix employed within its recommender systems. This provides the overall research paper more information and details on the inner workings of the recommender systems and what types of information is needed to reach a level of a ‘well-personalized’ product.

4. Project Design

Before beginning to tackle the goal set before us, the entire World Ready Team had a large team meeting to discuss the goals, expectations, and requirements of the entire project. Within this meeting, I along with the other interns, had a chance to ask questions and gain a better understanding of the problem at hand. We helped plan the workflow for the remainder of our 12-week internship.

4.1. System Requirements/Planning

First, we created a high-level pipeline of the project to map out our course. The entire pipeline can be seen below. These events required the collaboration of multiple teams for full functionality. The events highlighted with red borders were ones that my team were directly responsible for.

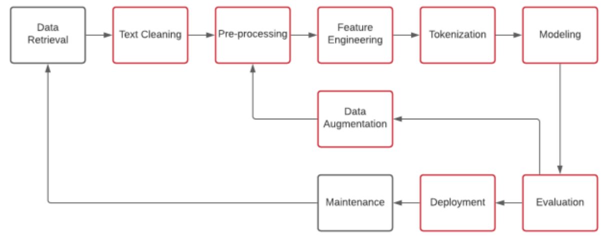

Figure 1. High-level Pipeline of the Personalization Project

Since we were assigned to deploy this project for the Polish market, some of our metrics to measure the effectiveness of our progress were qualitative as well as quantitative. We organized weekly meetings with Polish language experts both inside and outside of the company. These experts’ role was to verify and fill in any details regarding information lost in translation, synonym usage and overall semantics and diction. In addition, we planned on leveraging the industry standard metrics within the NLP and ML world, and much more specific metrics based on our choice of language model, encoders, and decoders.

Within the first several weeks, we had setup the workflow of the project as well as launched several other periphery tools to help us accomplish our goals. We relied on Jira, a proprietary issue tracking product, to keep track of any roadblocks along the way. Whenever someone faced a challenge, we created “tickets” that everyone in the team could view, with the goal of informing our teammates of our progress. Our primary method of communication was through Slack for simple and quick questions, while we relied on Microsoft Teams for larger time-consuming meetings.

Every day, within our daily stand-up meetings, we utilized these tools to help us chart our progress and direction. Our manager required a summary of the work that was accomplished previously and assigned more work accordingly. We used this time to discuss any prevalent issues anyone was facing, and check-in on one another.

4.2. Data Gathering

With the mantra of "Garbage In equals Garbage Out," the team began web-scraping several prominent Polish websites that contained large corpora of heavily diversified semantics. We made use of the current Polish data that was available on the company servers. These were data that users had willingly provided the company such as name, pronoun, title, gender, age, common people they interacted with and preferences within the website. All of this was aggregated into a new server for the Personalization Project. In total, the entire dataset was 572MB in size which is approximately 286,000 pages of textual data.

4.3. Pre-processing and preparation

The next step in the process was to begin cleaning and pre-processing the data. We relied heavily on the Natural Language Toolkit (NLTK) for this step. Firstly, we began by removing any alpha-numeric characters from the corpora. Second, we began the process of Tokenization, which splits strings of words into a list of words. Third, we removed all of the common Polish stop words within the text data. Fourth, we Lemmatize/Stem words within the data. Finally, we padded pronouns to avoid inputting any bias into our overall model.

4.4. Model Selection and Tuning

We collectively decided to utilize a Bidirectional Encoder Representations from Transformers (BERT), a language model created by researchers at Google AI language. It is a model that understands words and their relationships with one another based on their sequence, and peripheral words (Horev 2018). Due to the nature of the dataset, the language model took nearly 20 days to complete its training, using a 20-80 training, testing split; where 80% of the data was used for training the BERT model, and 20% was used to test the BERT model. However, to make better use of time, we had decided to run the model on a stratified sample of data to see the effects and results of the language model and apply changes as necessary. This was a pragmatic solution to keep the pace of the project consistent.

4.5. Testing and Evaluation and Re-testing

Once the BERT model had run several times on the 80% of the corpora, we evaluated the model for its various metrics such as F1-Score, model accuracy, AUC scores, as well as a confusion matrix to determine whether or not the model was able to perform at a par level while also being able to classify pronouns in the correct context. After the first several iterations of the testing the model, we concluded that the model was not able to grasp the context of the sentences we had inputted. We determined the next course of action was to diversify the data ourselves.

We were able to utilize several data-augmentation techniques to provide us with much more robust data. Techniques such as Synonym Replacement, Random Deletion, Random Swap, and Random Insertion when applied to a corpus are essentially able to create a foreign corpus identical in meaning but semantically different.

Once we had achieved an acceptable level of classification and language generation, we reported out results and methodology to the Polish language experts for them to verify whether our BERT model was ready for the professional world.

4.6. Packaging/Release

Finally, once we were satisfied with reaching an acceptable level of personalizing test sentences, we were able to move on to deployment of our project. We relied on a tool known as "Pickle" within the Pandas library on Python to help us deploy our BERT model. Picking the model condenses the information to better focus on effectively getting an output. We stored the models on the World Ready Team's GitHub repository for secure and quick access by anyone on the team.

I was personally in charge of developing an Application Programming Interface (API), able to communicate with any given consumer and the pickled BERT model to receive a personalized sentence. The API should be able to input a sentence, and user information, in Polish, and receive a personalized version of the same sentence all within a matter of seconds. I built the API using the company standard of MongoDB for the data storage and management and Node JS stack for the structure and aesthetics. I chose to implement a RESTful API architectural style to allow for the quick and secure form of query in addition to making it very lightweight and very feasible for scalability, especially for a company as large as PayPal. Furthermore, with the API creation, documentation also had to be created for any company employee to utilize the tools available.

5. Results

Currently the API that I had created is in use by several other teams in creating Chatbots to converse with customers, sprucing-up customer pages once logging in, as well as writing emails for clients regarding their PayPal accounts. The impact of the work that was carried out by the World Ready Team is something that can still be felt to this day. The work still lives through implementation from many other teams within the company. They have the potential and the ability to disseminate the tool we have built into many other facets for the company.

6. Conclusions

Throughout the process I was pushed outside of my comfort zone and surprisingly, was able to perform really well with the help of those around me. The tools that we had built within the 12-weeks are still being used to date and have grown substantially. The same API and BERT model that were made for the Polish market, has now been applied to the Spain market. Other Software Engineers were able to learn and analyze our software for the Polish market and apply the same tools we had used to create a similar project for many other markets. Ultimately, the company was able to see a rise in their original overall customer satisfaction of quality to 90%, as of February 2022. In addition, it has doubled the market share where the technology has been catered to.

7. Future Work

The work completed by the team during the summer was the initial step in a long process of applicability. Any implementation which relies on the NLP techniques that we have forged will be able to greatly succeed in achieving their goals. Currently, the team's objective shifted to horizontal scaling, and spreading the technology's use amongst different markets.

8. Acknowledgements

I would like to express my gratitude towards the team that I have spent my summer with, working, learning and thriving with. I am grateful to have spent my internship completing a practical tool which has a global reach. Special thanks to Nitin, John and Chang.

9. References

"Bert." *BERT*, https://huggingface.co/docs/transformers/model_doc/bert.

Flek, Lucie. "Returning the N to NLP: Towards Contextually Personalized Classification Models." *ACL Anthology*, Association for Computational Linguistics, July 2020, https://aclanthology.org/2020.acl-main.700.

Horev, Rani. "Bert Explained: State of the Art Language Model for NLP." *Medium*, Towards Data Science, 17 Nov. 2018, https://towardsdatascience.com/bert-explained-state-of-the-art-language-model-for-nlp-f8b21a9b6270.

Wei, Jason. "These Are the Easiest Data Augmentation Techniques in Natural Language Processing You Can Think of..." *Medium*, Towards Data Science, 11 Feb. 2019, https://towardsdatascience.com/these-are-the-easiest-data-augmentation-techniques-in-natural-language-processing-you-can-think-of-88e393fd610.