Undergraduate Thesis Prospectus

DFit: Cloud-Based Service and Business Case

(technical research project in Systems Engineering)

Technically Impoverished: The Homework Divide in America

(sociotechnical research project)

by

John Blair Mitchell III

October 27, 2023

technical project collaborators:

Franklin Glance
Esther Yi

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.

John Blair Mitchell III

Technical advisor: Roman Krzysztofowicz, Department of Systems and Information Engineering

STS advisor: Peter Norton, Department of Engineering and Society

General research problem

How may computational tools best support human decision-making?

Making an informed decision can be generalized into a few basic steps: determine the objective of the decision, gather and review relevant and available information, and make the decision that best fulfills the objective based on that information. The objective typically relies on human reasoning to determine, but the latter two steps have room to be automated, as computers can methodically process information exponentially faster than humans (Sharma, 2020).

Everyone makes decisions, so everyone stands to benefit from improving the process of doing so. However, with computers being a vital tool in this, it's important to recognize that their availability is not equal across the world. To support human decision-making as a whole, one must not just consider how to improve computational information processing, but also how to improve the availability of computers in the first place, as around 35% of the worlds' population still isn't connected to the internet (Petrosyan, 2023).

DFit: Cloud-based Service and Business Case

How can the performance of probabilistic distribution-fitting software be improved?

This is a Systems Engineering capstone project that I'm working on with Franklin Glance and Esther Yi, which is advised by Dr. Roman Krzysztofowicz. We're working on revitalizing DFit, a software for fitting data to probabilistic distributions. It was initially created by CapTech Consulting and is referenced in Dr. Krzysztofowicz's upcoming book, *Probabilistic Forecasts and Optimal Decisions* (Krzysztofowicz, 2024). This provides the main deadline for the project, as we need to have the software operational by the book's intended release in March 2024.

Distribution-fitting is a major part of any real-world forecasting problem, and those forecasting problems are becoming increasingly relevant as companies and researchers begin opting for stochastic models over simpler deterministic models. For this task, DFit offers a different methodology than most softwares on the market, allowing for more flexibility in the type of data that can be fit. It can handle random samples, censored samples, expert-assessed quantiles, and performs the fit using Dr. Krzysztofowicz's Uniform Distance method. Some may recommend machine learning techniques when this type of versatility is desired, but opting for a statistical method allows for much more explainability in the models that are created (Smeets, 2021).

The first part of this project, led by Franklin Glance, is to bring DFit up to modern standards. It's currently running on premises at CapTech's Richmond office, which leaves little room for scalability in preparation of the book's release. To resolve this, we'd like to move it to a cloud-based infrastructure using Amazon Web Services (AWS). We have the option of porting the current software to AWS, but it's built on antiquated technologies, so we would have to troubleshoot through getting them running on their modern, stable counterparts first. Instead, we're planning to completely overhaul the frontend web application. This will allow us to bring its design up to current standards for an easy deployment on the cloud, but leaves us essentially starting from scratch. However, the fitting engine on the backend can relatively easily be plugged into AWS Lambda, which will run the engine whenever a user requests a fit to be performed. Then, the fit results will be uploaded to a database that the new frontend can access to display them for the user.

Esther Yi is leading the next part of the project, which is to compare DFit to its potential competitors. These are other distribution fitting softwares like Stat::Fit and code libraries like R's

fitdistrplus. Our comparisons will be based on two types of analyses: distribution-fitting performance and software usability. For distribution-fitting performance, we'll compare the softwares' fits for samples with known distribution and parameters, as well as their fits for samples with unknown distribution and parameters. Most other softwares use more basic estimation methods, such as maximum likelihood, so our expectation is that we'll see DFit excel in this analysis. For software usability, we'll be concerned with the amount of time it takes to get a fit, the front-end design, the amount of data that can be handled, the amount of documentation, and the number of supported probability distributions, among other things. DFit's performance here will rely heavily on how well our implementation goes in the first part of the project. With these analyses, we can get a better understanding on where DFit will stand in the market following its release.

I'll be leading the last part of the project, which is to forecast the demand of the book and the software, as well as the revenue of the software, over the next three years. This is a challenging forecasting problem, as the book has many sources of demand across the sciences and engineering disciplines that it's applicable to, but is important nonetheless as CapTech should be able to justify the cost of the program. First, we will need to create conceptual models of the dynamics between the demand of the book, the demand of DFit, and the revenue generated by DFit. With the conceptual models built, we will carry out the methodology from the book to generate statistical forecasts. The first two portions of this project will become very important here, as whether or not the software is superior to its competitors will be a major factor in its demand, and especially its revenue.

If successful, by the end of the project, we will have three things: (1) DFit operating fully on the cloud with a modern design, (2) an analysis of DFit's performance relative to other

softwares on the market, and (3) three-year forecasts for the demand of *Probabilistic Forecasts and Optimal Decisions*, the demand for DFit, and the revenue for DFit (Krzysztofowicz, 2024).

Technically Impoverished: The Homework Divide in America

In the US, how are advocacies striving to diminish the homework divide?

In the United States, the digital divide imposes numerous inequities, but since the coronavirus pandemic, educational inequities have been the most prominent. Education systems are increasingly using technology such as computer applications and the internet, especially with remote instruction, but certain counties and students are being left behind. For example, a parent in Hartford, Connecticut, told NBC that some local students had to go to McDonalds to get Wi-Fi to complete their work (Mayes, 2020).

Technology maintenance theory says that most of the impoverished US has adopted digital technology, but their access is unstable and frequently interrupted. This makes it difficult for affected students to get better support, as they're already seen as "connected" from an outsider's perspective (Gonzales, 2016). Hernandez and Faith (2023) found that this was also a problem that frequently affected scientific surveys, as the choice for internet access was binary in most cases. The consequence of this is that affected individuals will be more skeptical to fully integrate technology into their lives, which is supported by a Chinese study that found that rural mobile users connect to networks with much more stable behavior than their urban counterparts (Li et al., 2015).

The Center on Reinventing Public Education is a research organization focused on improving the landscape of public schools by targeting four broad topics: students and families, systems, innovations, and politics of chance. They have a specific interest in the homework

divide, and they've shown that students who face the most obstacles with completing schoolwork at home are lower income and either rural or urban (Lake & Makori, 2020). This is the same group of students that's being targeted by EveryoneOn, which seeks to help connect them with low-cost internet and computers. This is because EveryoneOn found in their 2022 National Study that many low-income families can't afford even $100 for a computer, which is a main factor as to why 47% of survey respondents without internet access don't have internet access (Horrigan, 2022).

Governmental intervention is a major step in addressing these problems, with the Office of Educational Technology (OET) being one of the groups intended to help carry this out. Their mission is to promote digital equity, and an example of this is their highlighting of the impact of Washington's increased educational funding on rural school districts. This includes a grant that seeks to attain a 1:1 student to learning device ratio, which better prepares those students for a technological future (OET, 2023). The National Digital Inclusion Alliance also seeks to advance digital equity, in their case by supporting community programs and legislation. In fact, they even put their own executive director in front of Congress to advocate for improved broadband deployment across the country (Tesfaye, 2023). The last participant I'd like to mention, the Electric Frontier Foundation, has also been outspoken about governmental intervention in the space of digital equity. Liu (2023) criticized California's Middle Mile Network, which is a multibillion-dollar project that is supposed to bring broadband to low-income communities. Instead, it's seen revisions that are causing it to run counter to this goal, leaving out many of the communities that it was meant to support. This all makes it clear that closing the homework divide through promoting digital equity is a goal that many organizations, including governmental agencies, support.

References

Gonzales, A. (2016, February). The contemporary US digital divide: from initial access to technology maintenance. *Information, Communication, & Society 19(2)*, 234–248. Web of Science.

Hernandez, K., & Faith, B. (2023, June 27). Online but still falling behind: measuring barriers to internet use 'after access'. *Internet Policy Review 12(2)*. Web of Science.

Horrigan, J. (2022). The State of Digital Equity. EveryoneOn. https://www.everyoneon.org/2022-national-research

Krzysztofowicz, R. (2024). *Probabilistic Forecasts and Optimal Decisions*. John Wiley & Sons, Ltd.

Lake, R., & Makori, A. (2020, June). The Digital Divide Among Students During COVID-19: Who Has Access? Who Doesn't? Center on Reinventing Public Education. https://crpe.org/the-digital-divide-among-students-during-covid-19-who-has-access-who-doesnt/

Li, Y.; Jiang, H.; Yi, S.; Zhou, C.; Wu, L.; & Wu, M. (2015, October). Understanding Access Stability and Mobility Patterns of Mobile Internet Users. *2015 IEEE International Conference on Computer and Information*. 18-22. Web of Science.

Liu, C. (2023, September 8). California's Middle Mile Network Must Bridge the Digital Divide, Not Reinforce It. Electronic Frontier Foundation. https://www.eff.org/deeplinks/2023/09/californias-middle-mile-network-must-bridge-digital-divide-not-reinforce-it

Mayes, L. (2020, July 14). Parents and Teachers Discuss Digital Divide Hurting Students During Remote Learning. NBC Connecticut. https://www.nbcconnecticut.com/news/local/parents-and-teachers-discuss-digital-divide-hurting-students-during-remote-learning/2302394/

Office of Ed Tech. (2023, June 6). Digital Equity Champions for All Learners: Leaders Address Adoption Barriers to Advance Digital Equity. Medium. https://medium.com/@OfficeofEdTech/digital-equity-champions-for-all-learners-leaders-address-adoption-barriers-to-advance-digital-4bf168c47a2f

Petrosyan, A. (2023, October 25). Number of internet and social media users worldwide as of October 2023. Statista. https://www.statista.com/statistics/617136/digital-population-worldwide/

Sharma, K. (2020, April 20). How does the human brain compare to a computer? Crucial Blog. https://www.crucial.com/blog/technology/how-does-the-human-brain-compare-to-a-computer

Smeets, L. (2021, May 3). Machine learning is great, but is it going to crush statistical forecasting? OMP. https://omp.com/blog/machine-learning-is-great-but-is-it-going-to-crush-statistical-forecasting

Tesfaye, T. (2023, May 30). NDIA Advocates for Digital Equity Before House Subcommittee. National Digital Inclusion Alliance. https://www.digitalinclusion.org/blog/2023/05/09/ndia-advocates-for-digital-equity-before-house-subcommittee/