Ribbon Communications:

FLATR–Federated Learning Across

Telecommunication Resources

By Ved Nigam, Subre Moktar, Orlando Malanco, Kenan Stredic

Abstract

The purpose of the FLATR project is to be able to make a model using data from different clients, but also being able to privatize the data between companies. We will be using Federated Learning to aggregate models, rather than datasets. After having multiple client files running, we would like to aggregate the model, unpack the model, and use explainable AI to interpret the model. We would like to be able to identify significant features in a model to be able to determine the reason(s) for call quality from a Call Detail Record (CDR). Our deliverable will be a working architecture of Client-Server files and a Streamlit application to explain the aggregate model.

INTRODUCTION

The Federated Learning Across Telecommunication Resources (FLATR) project is an Analytics based dashboard to monitor any kind of data that Ribbon receives. The company receives data from IoT devices, 5G/4G/3G devices, OTT, Internet, voice, and any sort of communication records. The Analytics product that is part of the FLATR project will allow for data aggregation and enrichment from different companies that employ Ribbon services but will find a way to maintain data security within each company using Federated Learning. The broader idea is to simply share the Machine Learning models and the analytics from them instead of the data itself to maintain privacy in each company, but an aggregation of these models will allow for better informatics and incident management.

The result of this project would be to run the Ribbon Analytics product, which is going to gather data and run basic models, on the client's side; this model will be connected to a server on Ribbon's side, which will be processing information from the different companies. Our team's job is to identify a model that will get us the optimal results.

While we have metrics like accuracy and validation loss, Ribbon also requested an Explainable AI application from us. This is becoming a popular demand from governments in order to recognize the features that are determining decisions companies make using Machine Learning models. Our deliverable to Ribbon was a functioning Federated Learning Architecture using the Flower framework and an Explainable AI product on Streamlit.

RESOURCES

- Hardware:

  - Ribbon Laptops (Windows OS with Intel Chips)
- Software:
  - PuTTy
  - WinsCP
  - Linux/Bash
  - Python
  - Jupyter Notebook
  - VSCode
- Python Libraries:
  - Numpy
  - Pandas
  - Datetime
  - Streamlit
  - Seaborn
  - Matplotlib
  - Shap (Shapley)
  - Pytorch
  - Flower

KEY ROLES

All UTD group members worked as equals. When communication was needed, we would quickly call and one of the members would email the Company Mentor with all of us copied on the email. If calling to communicate was not an option, we also had a Text and Discord chat. The

Company Mentor sent out the outline for our weekly meeting after the meeting, so we used those as our notes for the rest of the week. Email communication was maintained with the mentor throughout the week.

COMMUNICATION PLAN

The teammates interacted through Text, Discord, and MS Teams calls. Text and Discord would be used for general communication and setting up meetings between the UTD members. Teams was the official form of communication for video chats and checkpoint calls with Ribbon Mentors. Company emails were the main form of communication with the Ribbon mentors outside of our weekly checkpoint calls.

TIMETABLE

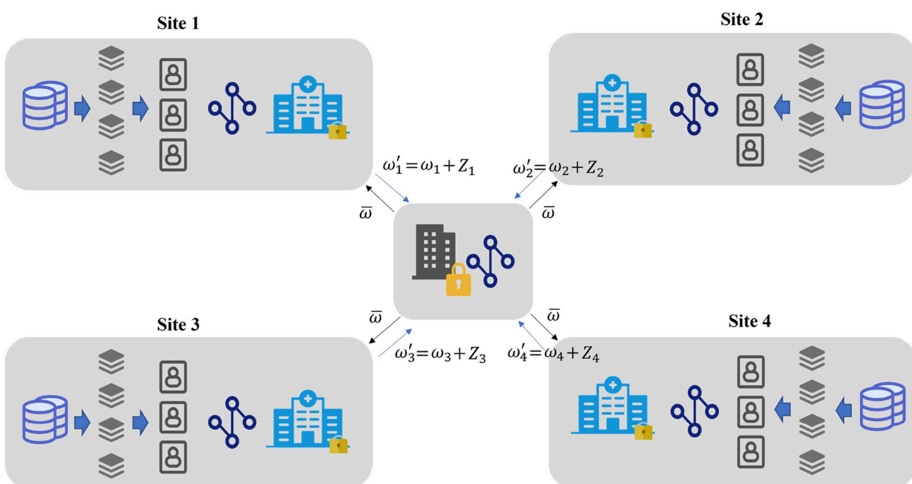
The final product that Ribbon expects from us is a functioning Federated Learning architecture which is somewhat informative of the data. It does not need to have high accuracy, But, our work should highlight some of the features that are important in the data so that we can learn from it. In addition to the Federated Learning architecture, this will be handled through an Explainable AI application on Streamlit.

EVALUATION

The most prevalent indicator that the project was successful was the reaction of our mentors. We were able to work on all parts of the project that the mentors expected from us. The Federated Learning architecture is able to be deployed on different kinds of linux machines for data collection and querying. Elaborating more on the federated Learning architecture, we included aggregation strategies for models that Ribbon can pick from. We also have an Streamlit application which implements Explainable AI allowing us to give interpretable information about black box models such as Neural Networks and Deep Learning models.

IMPLEMENTATION DETAILS

While all the code and design is protected under the NDA we signed with Ribbon, attached below is a general diagram of the Federated Learning architecture.



The image shows a very general overview of how multiple clients will connect to one server. The formula that we see on the server entity represents the aggregation strategy previously mentioned. Each client (different machines) will run the same model file, and the weights/parameters will be shared to the server. The server will combine all provided models using a strategy (Mean, Median, Weighted Average); this aggregated model will be saved on the server side.

Once the model is saved, our Explainable AI application will be able to read this model and identify key features in the model. This evaluation will only be accessible to Ribbon for privacy. The problem is that the data for the explainable AI needs the a sample of the data to get the explanations.

ISSUES AND LESSON LEARNED

As with any project, we encountered several challenges that required careful navigation and problem-solving throughout our journey with the FLATR project. These challenges ranged from domain-specific knowledge gaps to technical hurdles in accessing and utilizing essential tools. In this section, we delve into the specific issues we faced and the strategies employed to overcome them.

1. **Limited Expertise in Telecommunications Realm**: One of the primary challenges we encountered during the project was our limited expertise in the telecommunications domain. As data science students, we initially lacked comprehensive knowledge of the specific nuances and intricacies of telecommunications data. This posed a significant hurdle in understanding the context and requirements of the FLATR project.
2. **Adapting to Accessing Necessary Tools on Provided Laptops**: Another obstacle we faced early in the project was the challenge of adapting to accessing necessary tools on the laptops provided to us. As the project required us to work with specific software and tools for data analysis and model development, navigating and configuring these tools initially proved to be challenging.
3. **We needed the data for the streamlit app:** We use federated learning to keep the data in their secure places, although for the streamlit app for explainable AI we need the data or sample of the data. Therefore we need to figure out a way where the clients can use the

streamlit app on their systems. Again, to make sure that we keep the data privacy the same.

Through proactive learning, collaboration, and perseverance, we successfully navigated through these challenges. Our experience in addressing domain-specific knowledge gaps and technical hurdles not only enhanced our skills but also reinforced the importance of adaptability and resilience in project work.

Our journey with the FLATR project has been a significant learning experience, enriching our capabilities as data scientists. We acquired proficiency in various programming languages and tools essential for data science, such as Python, SQL, and Linux/Bash. Additionally, our familiarity with development environments like Visual Studio Code, Jupyter Notebook, and Streamlit deepened, enabling us to work efficiently on data science projects.

Exploring Federated Learning machine learning techniques involved substantial coding, research paper analysis, and online research, providing us with valuable insights and skills for our future careers in data science. Furthermore, the project facilitated hands-on experience in accessing data from servers and working collaboratively, enhancing our teamwork skills and preparing us for future collaborative endeavors.

Throughout the project, the invaluable guidance and support from our mentors played a pivotal role in our learning journey, providing us with tools and strategies for success. As a result, we have strengthened our skill set and gained confidence in tackling complex data science challenges in future projects.

ETHICS DISCUSSION

The implementation of Federated Learning in the FLATR project raises important ethical considerations regarding data privacy, security, and transparency.

1. **Data Privacy and Security**: One of the primary ethical concerns addressed by FLATR is the protection of user data privacy. By adopting Federated Learning, the project ensures that sensitive data from different companies remains decentralized and never leaves their respective premises. This approach mitigates the risk of data breaches and unauthorized access to personal information, aligning with principles of privacy protection.
2. **Transparency and Accountability**: FLATR emphasizes transparency in its approach to data aggregation and model training. Through Explainable AI applications, the project enables stakeholders, including Ribbon and its clients, to understand the decision-making process of the machine learning models. This transparency fosters accountability and trust among users, ensuring that model predictions are based on fair and interpretable criteria.
3. **Fairness and Bias Mitigation**: An important aspect of ethical AI is addressing biases in the data and ensuring fairness in model outcomes. FLATR should incorporate mechanisms to identify and mitigate biases during model training and evaluation. Additionally, ongoing monitoring and evaluation of model performance can help detect and rectify any instances of unfair treatment or discrimination.

4. **Informed Consent and User Empowerment**: FLATR should uphold principles of informed consent, ensuring that users are aware of how their data is being used and for what purposes. Providing users with control over their data and the option to opt out of data sharing initiatives promotes autonomy and empowers individuals to make informed decisions about their privacy.
5. **Collaborative Governance and Regulation**: Ethical AI initiatives like FLATR should operate within a framework of collaborative governance involving stakeholders from various sectors, including government, industry, academia, and civil society. Additionally, adherence to relevant regulations and standards, such as GDPR in Europe or CCPA in the United States, ensures legal compliance and reinforces ethical practices.

By addressing these ethical considerations, FLATR not only facilitates the development of innovative analytics solutions but also promotes responsible and ethical use of data in telecommunications and beyond.

CONCLUSION

This project is important to analyze the communication between large corporations and the general population without having to share such massive and private records regarding each communication. By applying federated learning in this way, Ribbon will be opening an avenue to be able to deliver a better product that is more fault tolerant.

FUTURE WORK

In continuing our efforts to refine and expand the FLATR project, several areas have been identified for future development. Building upon the foundation established in our presentation, we aim to further enhance the capabilities and effectiveness of the FLATR system through the following initiatives:

1. **Expand Streamlit App Functionalities**:
    a. **Incorporate Additional Model Evaluation Metrics**: Enhance the Streamlit application to include a wider range of model evaluation metrics beyond accuracy and validation loss. Metrics such as precision, recall, F1 score, and ROC curves can provide a more comprehensive understanding of model performance.
2. **Explore Federated Learning Strategies**: Experiment with different federated learning strategies to optimize model aggregation and training across multiple clients. This could involve investigating techniques such as differential privacy, secure aggregation, and adaptive learning rates to improve the efficiency and effectiveness of the federated learning process.
3. **Integrate Models from Alternative Datasets**: Extend the Streamlit application to support the integration of models trained on alternative datasets. By incorporating diverse data sources, the application can provide more robust and adaptable analytics capabilities to address a broader range of use cases and scenarios.
4. **Augment Data Visualization Options**: Enhance the data visualization capabilities of the Streamlit application to provide more detailed insights into specific call record error codes. Implement interactive plots and visualizations tailored to highlight patterns and

trends associated with different error codes, enabling users to quickly identify and address issues.

5. **Implement LIME for Explainable AI**:
   a. **Assess its Performance Compared to SHAP**: Integrate the Local Interpretable Model-agnostic Explanations (LIME) technique into the Streamlit application for explainable AI. Evaluate its performance and effectiveness in providing interpretable explanations for model predictions, comparing it with the SHAP (SHapley Additive exPlanations) method. Conduct thorough testing and validation to determine the suitability of LIME for the project's specific requirements and objectives.
6. **Aim to Run All Three Clients on the Analytics Tool**:
   a. **Server Code Execution on the Playground Server**: Work towards running all three clients (different machines) on the analytics tool, with server code execution managed on the playground server. Streamline the deployment and execution process to ensure seamless integration and interoperability between the clients and the server. Implement robust communication protocols and error handling mechanisms to facilitate reliable data exchange and collaboration among the distributed components.
7. **Move the Streamlit Server to be on Client side**
   a. The problem is for shap values we need the data and for more understanding we need to access the data to label columns. Since this is a problem because we wanted federated learning to be used for data privacy. We might send the streamlit

files to be used in their servers so that we can use shap value more properly and in conjunction with their data.

By pursuing these future works, our team can further enhance the functionality, performance, and usability of the FLATR project, advancing its capabilities in federated learning, explainable AI, and analytics for telecommunications data.

CONTACT INFORMATION

Team Members:

Name: Kenan Stredic

Email: kstredic02@hotmail.com

Linkedin: https://www.linkedin.com/in/kenanstredic/

Github: https://github.com/KenanStredic

Name: Subre Moktar

Email: subremoktar@gmail.com

Linkedin: https://www.linkedin.com/in/subre-moktar/

Github: https://github.com/Subre-Moktar

Name: Ved Nigam

Email: ved.nigam@icloud.com

Linkedin: https://www.linkedin.com/in/ved-nigam-b81b891b6/

Github: https://github.com/ved-n/

Name: Orlando Malanco

Email: Orlandomalanco@gmail.com

Linkedin:https://www.linkedin.com/in/orlando-malanco-25949524b/

Github:

Ribbon Communications Mentors:

Name: Patrick Sollee

Email: psollee@rbbn.com

Name: Chip Boyle

Email: cboyle@rbbn.com

Name: PS Ravikiran

Email: pravikiran@rbbn.com

Faculty Mentor:

Name: Dr. Octavious Smiley

Email: Octavious.Smiley@UTDallas.edu

REFERENCES

The sources utilized throughout our project include academic papers, textbooks, online resources, datasets, software tools, and mentor guidance. These materials have been instrumental in our research, development, and implementation processes.

Academic Papers:

1. Yang, et al. "Federated Machine Learning: Concept and Applications," arXiv:1902.04885v1, 13 Feb 2019.
2. Lim, et al. “Federated learning in mobile edge networks: A comprehensive survey,” arXiv:1909.11875v2 [cs.NI] 28 Feb 2020.
3. Niknam, et al. "Federated Learning for Wireless Communications: Motivation, Opportunities and Challenges," arXiv:1908.06847v4, 3 May 2020.
4. Li, et al. "A Survey on Federated Learning Systems: Vision, Hype and Reality for Data Privacy and Protection,” arXiv:1907.09693v7 [cs.LG], 5 Dec 2021.
5. Brecko, et al. "Federated Learning for Edge Computing: A Survey," Appl. Sci. 2022, 12, 9124.

Software Tools and Platforms:

- Flower: https://flower.dev
- Flower Documentation: https://flower.dev/docs/
- Flower Example Walkthrough (PyTorch MNIST): https://flower.dev/docs/examplewalkthrough-pytorch-mnist.html

Additional Resources:

- Shap Documentation: https://shap.readthedocs.io/en/latest/example_notebooks/overviews/An%20introduction%20to%20explainabl
- Lime Documentation: https://lime-ml.readthedocs.io/en/latest/lime.html

Access to all project code, READMEs, and pertinent files can be found on the Ribbon Communications Microsoft Teams platform within the Ribbon Labs/00_Federated Learning/FL_Projects folder, under FLATR_SPR24.

Ved Nigam | Subre Moktar

Print Name Team Member 1 | *Print Name Team Member 2*

Kenan Stredic | **Orlando Malanco**

Print Name Team Member 3 | *Print Name Team Member 4*

Pat Sollee

Company Mentor Name, Company Mentor

DocuSigned by:

Faculty Advisor Name, Faculty Advisor

DocuSigned by:

Charles Boyle

Company Mentor Name, Company Mentor