FEDERATED LEARNING-BASED BIG DATA FOR EDUCATION SYSTEM

Project Submitted to the
SRM University AP, Andhra Pradesh
for the partial fulfillment of the requirements to award the degree of

Bachelor of Technology
in
Computer Science & Engineering
School of Engineering & Sciences

submitted by

Pallavi Annapareddy(AP20110010564)

Jayabhavani Angajala(AP20110010567)

Gayathri Kondaveeti(AP20110010568)
Meghana Amara(AP201100100577)

Under Guidance of

Prof. Sriramulu Bojjagani

Department of Computer Science & Engineering
SRM University-AP
Neerukonda, Mangalgiri, Guntur
Andhra Pradesh - 522 240

May 2024

DECLARATION

I undersigned hereby declare that the project report FEDERATED LEARNING-BASED BIG DATA FOR EDUCATION SYSTEM submitted for partial fulfillment of the requirements for the award of degree of Bachelor of Technology in the Computer Science & Engineering, SRM University-AP, is a bonafide work done by me under supervision of Prof. Sriramulu Bojjagani. This submission represents my ideas in my own words and where ideas or words of others have been included, I have adequately and accurately cited and referenced the original sources. I also declare that I have adhered to ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the institute and/or the University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of any degree of any other University.

Place	:	Mangalagiri	Date	:	May 12, 2024
Name of student	:	Pallavi Annapareddy	Signature	:	....
Name of student	:	Jayabhavani Angajala	Signature	:	....
Name of student	:	Gayathri Kondaveeti	Signature	:	....
Name of student	:	Meghana Amara	Signature	:	....

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

SRM University-AP

Neerukonda, Mangalgiri, Guntur

Andhra Pradesh - 522 240

CERTIFICATE

This is to certify that the report entitled FEDERATED LEARNING-BASED BIG DATA FOR EDUCATION SYSTEM submitted by **Pallavi Annapareddy, Jayabhavani Angajala, Gayathri Kondaveeti, Meghana Amara** to the SRM University-AP in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in computer science and engineering is a bonafide record of the project work carried out under my guidance and supervision. This report in any form has not been submitted to any other University or Institute for any purpose.

Project Guide | Head of Department

Name :Prof. Sriramulu Bojjagani	Name :Prof. Niraj Upadhayaya
Signature:	Signature: ....

ACKNOWLEDGMENT

I wish to record my indebtedness and thankfulness to all who helped me prepare this Project Report titled FEDERATED LEARNING-BASED BIG DATA FOR EDUCATION SYSTEM and present it satisfactorily.

I am especially thankful for my guide and supervisor Prof. Sriramulu Bojjagani in the Department of Computer Science & Engineering for giving me valuable suggestions and critical inputs in the preparation of this report. I am also thankful to Prof. HoD Name, Head of Department of Computer Science & Engineering for encouragement.

My friends in my class have always been helpful and I am grateful to them for patiently listening to my presentations on my work related to the Project.

Pallavi Annapareddy, Jayabhavani Angajala, Gayathri Kondaveeti,
Meghana Amara
(Reg. No. AP20110010564, AP20110010567, AP20110010568,
AP201100100577)
B. Tech.
Department of Computer Science & Engineering
SRM University-AP

ABSTRACT

This paper proposes a novel approach to enhancing education systems through the integration of federated learning techniques with big data analytics. Traditional methods of data analysis in educational settings often face challenges regarding data privacy, security, and scalability. Federated learning addresses these issues by enabling collaborative model training across distributed datasets without the need for data centralization, thus preserving the privacy of sensitive information. By harnessing the vast amounts of educational data generated from various sources such as online learning platforms, student information systems, and educational applications, federated learning empowers educational institutions to derive valuable insights while respecting data privacy regulations. Leveraging the collective intelligence of decentralized data sources, federated learning algorithms facilitate the development of robust predictive models for student performance, personalized learning recommendations, and early intervention strategies. Moreover, federated learning enables continuous model improvement through the iterative aggregation of local model updates from participating institutions, ensuring adaptability to evolving educational landscapes. This paper explores the technical foundations of federated learning, its application in education systems, and the potential benefits it offers in terms of improving learning outcomes and fostering data-driven decision-making in education. Through a comprehensive review of existing literature and case studies, this research aims to provide insights into the opportunities and challenges associated with implementing federated learning-based big data analytics in education systems, ultimately paving the way for a more efficient and personalized approach to education.

CONTENTS

LIST OF FIGURES

Chapter 1

INTRODUCTION TO THE PROJECT

Federated learning, without exchanging raw data, enables several institutions to cooperatively train machine learning models on decentralised data. In addition to protecting data privacy, this strategy helps organisations to use collective intelligence to improve model performance. In our research paper we use the Gated recurrent unit(GRU) model to find the drop out rate of students in educational institutions within the federated center. Sending data to a server from clients in federated learning isn't about securing the data; rather, it's about merging insights from several sources to enhance learning models while maintaining privacy. Using their own data, each client trains a model, and only the insights gained.Raw data is not shared with a central server. The server builds a more comprehensive and accurate model by combining these insights, all while protecting the privacy of individual user data. Better forecasts and recommendations are made possible by this cooperative strategy, which also guarantees that private data is kept secure on clients' devices. Within the realm of education, federated learning-based data analytics holds significant promise for revolutionary change. By utilising the pooled intelligence present in federated institutions, it enables educational stakeholders to propel innovation in education, enhance student outcomes, and cultivate individualised learning experiences. Federated learning-based data analytics is the key to releasing the full potential of educational data, from resource optimisation and predictive analytics to personalised interventions and adaptive learning paths. In spite of the many benefits that federated learning offers, it is critical to recognise and take into account the related issues and concerns. The effective application of federated learning-based data analytics in education necessitates careful planning, stakeholder participation, and a strong technology infrastructure, much like with any transformational endeavour. Furthermore, using federated learning frameworks raises serious problems about data security, privacy, and ethical use. In considering this, our study aims to investigate how federated learning-based data analytics may transform the educational landscape. By creating a federated continual learning framework based on gated recurrent units (GRUs), we hope to tackle important issues and possibilities related to using distributed data to enhance education.

Chapter 2

MOTIVATION

2.1 REASONS WHY FINAL YEAR ENGINEERING PROJECTS ARE IMPORTANT

In the context of our project on Federated Learning for educational data analysis, final year engineering projects hold paramount importance for several reasons. Firstly, inspired by the innovative applications of Federated Learning in various domains, our project aims to address real-time challenges faced in educational institutions, such as predicting student dropout rates and enhancing educational outcomes. While the inspiration for our project may stem from existing research or industry trends, our faculties recognize the significance of such endeavors in preparing us for future career opportunities and endeavors.

By undertaking this project, we strengthen our core skills in machine learning, data analysis, and problem-solving. Through practical exposure to Federated Learning techniques and methodologies, we enhance our ability to tackle complex problems and make informed decisions. This hands-on experience not only bridges the gap between theory-based learning and practical applications but also equips us with the necessary skills to face future challenges in the engineering industry.

Furthermore, our project synthesizes the knowledge acquired throughout our academic years, demonstrating our aptitude in applying theoretical concepts to real-world problems. By working on multidisciplinary projects like Federated Learning for educational data analysis, we engage in group discussions, adapt to technological advances, and gain practical insights into the industry landscape. This not only enriches our academic experience but also prepares us for the demands of the engineering profession.

In addition, final year engineering projects offer us the opportunity to explore our areas of interest and contribute to research-based, industry-oriented projects. By delving into innovative domains like Federated Learning, we add value to our resumes and distinguish ourselves as aspiring engineers with a passion for cutting-edge technologies.

Our project on Federated Learning exemplifies the importance of final year engineering projects in providing practical exposure, enhancing core skills, and addressing real-world challenges. Through this endeavor, we not only fulfill academic requirements but also prepare ourselves for future endeavors in the engineering industry. Let's examine a few points that illustrate the significance of senior engineering projects, which are as follows:

2.1.1 It helps to identify a real-time problem and provide a solution

Final year engineering projects, such as our Federated Learning-based project, play a crucial role in addressing real-time challenges in the field of education. By leveraging federated learning techniques, we aim to analyze and predict the dropout rates of students in educational institutions within a decentralized data framework. This project allows us to tackle a pressing issue faced by educational stakeholders and provide innovative solutions that can improve student retention and success rates.

2.1.2 It helps to choose diversified research topics.

Diversification of research topics is essential for exploring the breadth and depth of engineering domains. In our project, we delve into the intersection of machine learning, privacy-preserving techniques, and educational data analysis. By choosing a diversified research topic like Federated Learning for educational data analytics, we expand our understanding of emerging technologies and their applications in addressing complex societal challenges.

2.1.3 It helps to choose appropriate project topics and mentor carefully.

Selecting an appropriate project topic and securing guidance from experienced mentors are critical aspects of project development. In our Federated Learning project, we carefully chose a topic that aligns with our interests in machine learning and education technology. Additionally, we benefit from the mentorship of experts in the field who provide valuable insights and guidance throughout the project, ensuring its success and impact.

2.1.4 Understand and analyze project documentation effectively.

Effective documentation is vital for capturing the essence of our Federated Learning project and communicating its findings to stakeholders. We meticulously document our research methodology, data analysis techniques, and experimental results to ensure clarity and reproducibility. By analyzing project documentation effectively, we gain a deeper understanding of the intricacies of Federated Learning and its implications for educational data analytics.

2.1.5 Effective planning

Effective planning is essential for the successful execution of our Federated Learning project. We carefully outline project goals, milestones, and timelines to ensure that we stay on track and deliver quality outcomes. Through strategic planning, we manage resources efficiently and anticipate potential challenges, enabling us to overcome obstacles and achieve our objectives effectively.

2.1.6 Provides a platform for self-expression

Our Federated Learning project provides a platform for self-expression, allowing us to showcase our creativity, problem-solving skills, and technical expertise. Through this project, we explore innovative solutions to address real-world challenges in education, demonstrating our ability to make meaningful contributions to the field. By expressing ourselves through our work, we inspire others and contribute to the advancement of knowledge and technology in engineering.Final year engineering projects play a vital role in shaping the future of engineering by providing students with opportunities to apply their knowledge, develop their skills, and make a positive impact on the world around them. Through rigorous research, innovative thinking, and effective collaboration, students emerge from their final year projects as confident, capable, and inspired engineers ready to tackle the challenges of the industry head-on.

Chapter 3

LITERATURE SURVEY

3.1 BACKGROUND

Federated Learning (FL) offers a novel method of analysing educational data that preserves institutional autonomy and data privacy while enabling cooperative model training. Through the decentralisation of the training process and the localization of data, FL allows academic institutions to participate in the building of models without exchanging confidential data. By conducting our research, we hope to use FL's potential to advance knowledge and advancements in the field of education, opening the door to data-driven decision-making that upholds individual autonomy and privacy. Our research focuses on using GRU (Gated Recurrent Unit) models to forecast dropout rates in federated centre educational institutions. The FL framework's application of GRU models makes it possible to forecast student attrition while maintaining data autonomy and privacy. Our goal is to give educational stakeholders useful information to boost educational outcomes and increase student retention by utilising FL and GRU models. By leveraging FL techniques, educational institutions can enhance predictive analytics capabilities while safeguarding privacy, paving the way for transformative advancements in education data analysis.

3.2 LITERATURE REVIEW

Federated Learning (FL) has emerged as a promising paradigm for collaborative model training across decentralized devices while preserving data privacy and security. This section provides an overview of FL and its significance in revolutionizing machine learning in distributed environments.

Foundational Work by Konečný et al.: Konečný et al.'s seminal work laid the groundwork for understanding the challenges and opportunities in FL. Their research highlighted the importance of tailored optimization algorithms to address communication constraints between the server and client devices.

Exploring Challenges of Non-IID Data Distribution by Zhao et al. : emphasized the challenge of non-Independent and Identically Distributed (IID) data across client devices in FL settings. Strategies to mitigate the effects of non-IID data, such as advanced aggregation methods and data augmentation techniques, were explored.

Applications in Education Brisimi et al. : Brisimi et al. focused on applying FL to solve binary supervised classification problems in the healthcare domain. Their work aimed to develop decentralized optimization frameworks for predictive analytics while preserving data privacy, serving as a model for FL applications in education.

Efforts to Reduce Communication Overhead Wang et al. : emphasized the importance of efficient communication protocols and model compression techniques to minimize data transmission during FL rounds. Their research aimed to optimize communication overhead, a critical aspect of FL implementation.

Security Considerations in FL by Yang et al. : delved into the security vulnerabilities of FL, including model poisoning attacks and inference attacks, threatening the privacy and integrity of FL systems. Their work underscored the importance of robust security measures in FL deployments.

Addressing Practical Implementation Challenges Kim et al. and Mohassel and Zhang : addressed practical implementation challenges in FL, including device heterogeneity and synchronization issues. Their research focused on ensuring consistent model performance across diverse devices and network conditions.

Advancements in FL for IoT Applications: Khan et al. : Khan et al. comprehensively reviewed recent advancements in FL, particularly in the context of Internet of Things (IoT) applications. Their work outlined metrics for evaluating FL advancements and devised a taxonomy for FL over IoT networks, providing valuable insights for researchers and practitioners.

Classification and Clustering of Literature Progress: Banabilah et al.: Banabilah et al. provided a classification and clustering of literature progress in FL across various application technologies and domains, including healthcare, education, and industry. Their work discussed future research directions and challenges, serving as a comprehensive reference for FL applications.

Chapter 4

DESIGN AND METHODOLOGY

4.1 ALGORITHM

Gated recurrent unit Gated Recurrent Unit (GRU) architecture is used in Federated Learning-Based Big Data for Education System. Recurrent neural network (RNN) architectures such as GRU are especially useful for modelling sequential data, which makes them highly relevant for tasks like temporal pattern analysis and educational interaction analysis within educational datasets. The Gated Recurrent Unit (GRU) architecture is essential for processing sequential data in a Federated Learning (FL) system designed for educational data analysis while maintaining user privacy. Let's investigate how the GRU gates might be modified to meet FL's demands in the field of education.:

1. Input Gate: When it comes to FL for education, the GRU unit's input gate chooses which fresh data from each educational institution's local dataset to include in the model. The flow of pertinent variables, including student demographics, academic standing, and learning habits, that are taken from the educational data is regulated by the input gate. By enabling only pertinent features to update the model parameters, it makes sure that sensitive data, such personally identifying information, is kept safe.

2. Update Gate :The amount of the prior model state that should be updated or maintained using fresh data from the local datasets is controlled

by the update gate in the GRU unit.When it comes to FL for education, the update gate allows the model to remain consistent over several decentralised training rounds while accommodating changing student and school factors. Incorporating new insights from the current training cycle and holding onto knowledge from earlier rounds is balanced.

3. Reset Gate : The GRU's reset gate uses the current input to determine how much of the prior concealed state should be forgotten or reset. The reset gate in FL for Education aids in the model's ability to adjust to variations in student populations, updates to curricula, or modifications in teaching approaches used by various educational establishments. As a result, the impact of out-of-date knowledge from earlier training cycles is reduced and new patterns and trends in educational data are captured by the model.

4. Hidden State Update : Once the input, update, and reset gates have calculated their respective contributions, the GRU unit's hidden state is modified. This updated concealed state contains the relevant information retrieved from the local datasets of participating educational institutions while adhering to privacy limitations. The hidden state is a simplified representation of the collective knowledge acquired through distributed training, allowing for the creation of insights and predictions about educational outcomes.

5. Output : The GRU unit's output, obtained from the final hidden state, can be used to perform a variety of educational analytics tasks, including predicting student performance, assessing dropout risk, and providing personalized learning recommendations. In FL for education, the output can be pooled across numerous participating schools to get global insights while keeping individual student data safe and decentralised. GRU gates preserve

privacy and security standards while enabling cooperative model training over dispersed datasets in a FL context for education. The GRU architecture facilitates the creation of reliable and privacy-preserving predictive models for educational analytics by selectively processing and updating data from each educational institution's local dataset.

4.2 METHODOLOGY

GRU Architecture(Components of GRU):

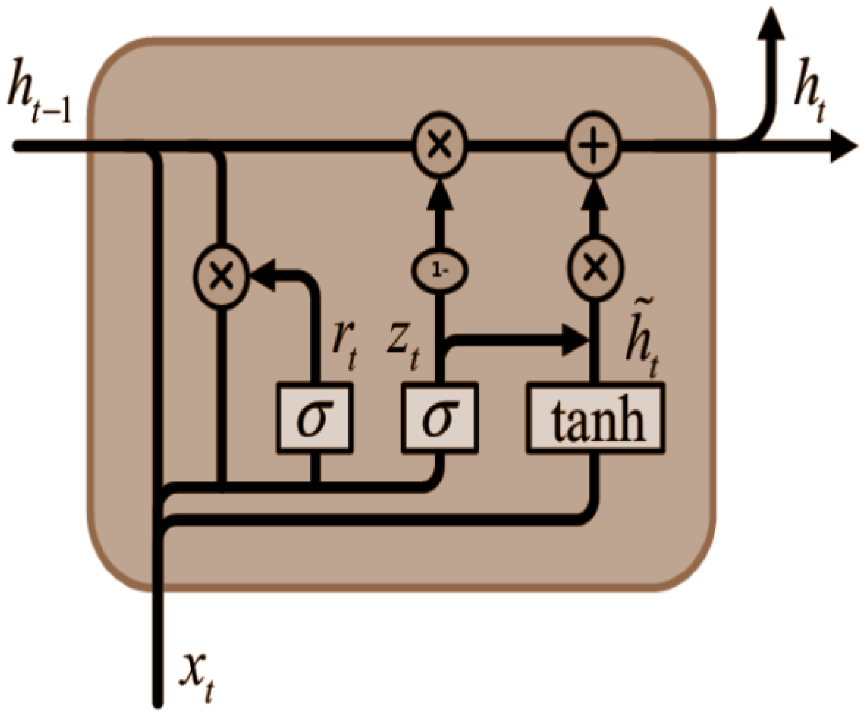

Figure 4.1: GRU Architecture

1.Update Gate (zt) and Reset Gate (rt): These gates control the flow of information at each time step.zt and rt are calculated as follows: zt=(Wz[ht1,xt]) rt=(Wr[ht1,xt]) where xt is the input at ime t, Wz and Wr are the weight matrices for the update and reset gates respectively, and ht1 is the output of the previous hidden layer.

2. Candidate Hidden State (h t):The candidate hidden state captures the information to be updated in the current time step.It is computed using the previous hidden state (ht1) and the current input (xt) with the reset gate (rt) controlling the retention of historical information: h t =tanh(W[rt ht1,xt]) where W is the weight matrix.

3. Hidden State Update (ht): The final hidden state (ht) is a combination of the previous hidden state (ht1) and the candidate hidden state (h t), controlled by the update gate (zt): ht=ztht1+(1zt)h t

4. Output:The output of the GRU (ylast) is calculated based on the final hidden state and is used for prediction or further processing: ylast=(Wohlast+bo) where Wo is the weight matrix of the output layer and bo is the bias vector. Training and Optimization: During training, GRU uses the backpropagation through time (BPTT) algorithm for network optimization.The loss function to optimize is the cross-entropy loss, which compares the predicted labels (ylast) with the true labels (y(i)) for each sample i. In summary, GRU effectively combines historical information with current inputs to predict future information, making it well-suited for tasks involving sequential data processing. Its gated architecture allows it to selectively retain or discard information, enabling efficient learning and representation of temporal patterns.

4.3 MODEL ARCHITECTURE

This model architecture is about how our model works using gru architecture.

Federated system architecture - The educational institutions such as schools or colleges(client systems) are connected to the main federated server system. In order to get the trained model update from the client system, there are 3 main steps.

Dataset: Each educational system(client) has its own dataset based on their own data,which contains different kinds of data about students.From its own databases or systems, each educational institution gathers pertinent data, including demographic data, learning behaviour data, and student performance records.

Preprocessing: In a dataset we may have various kinds of data such as textual data, noise, null values, redundant data etc. . . so that kind of data may not be accepted by our gru model, so we apply preprocess methods to convert the data into gru acceptable data .so data filtering,data anonymization,data encoding, data encryption are done.These pretreatment procedures ensure that the data is clean, anonymised, and appropriate for training machine learning models while protecting student privacy. This helps get the data ready for federated learning in the educational system.

Model: We use gated recurrent unit model, so the processed data to gru and it will train the dataset and give us an update. Because of its ability to analyse sequential data, such as student interactions or learning progressions, GRU (Gated Recurrent Unit) is essential to federated learning for education. It is skilled at comprehending the sequence and timing of events, which is important in educational environments, because of its capacity to capture temporal dependencies. In federated contexts with limited resources, GRU's computational efficiency is useful, and its local training capacity protects student privacy. Federated learning can produce a global model that enhances personalised learning by customising recommendations and interventions depending on individual learning paths, thereby improving the quality of education received. This is achieved by combining locally learned GRU models.

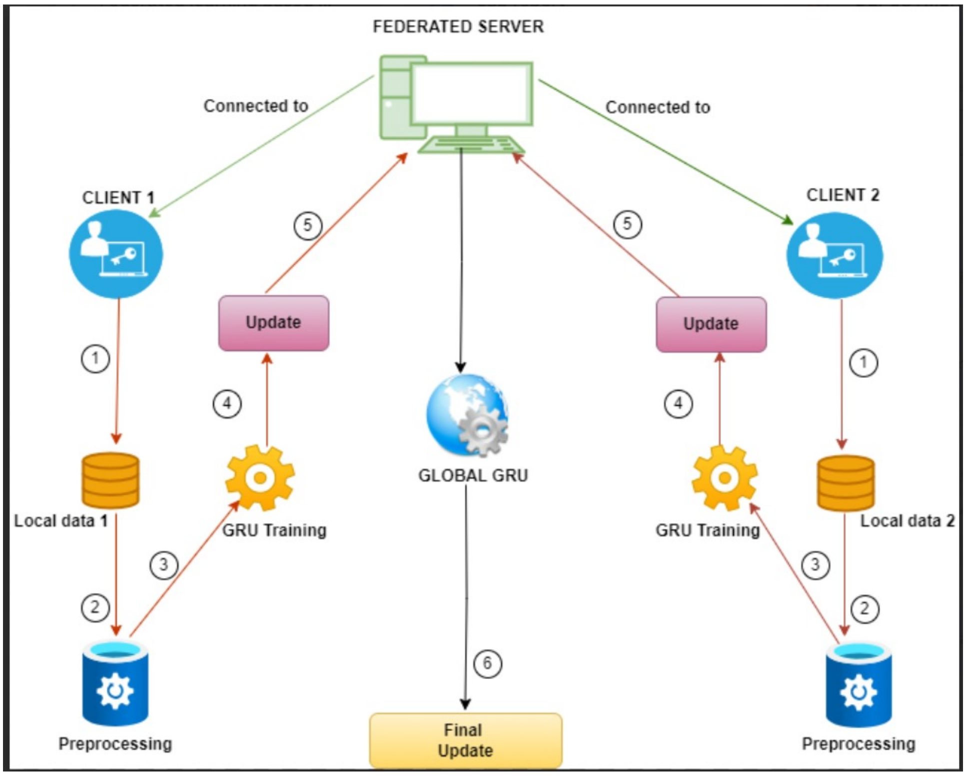

Figure 4.2: Server Client Architecture

Server Client representation - In a federated learning system, every client has a local GRU model and every federated server has its global GRU model. Input, hidden, and output layers typically make up the architecture of a GRU model, with GRU layers handling sequential data. A subset of the educational data is held by each of the several clients that receive it. Clients use their own data to train their own GRU models, maintaining the privacy of the model parameters. After being locally computed using the data from each client, model updates (gradients) are shared with a central server. Federated averaging, or some other aggregation technique, is used by the central server to compile the model updates from each client. The global parameters of the GRU model are updated via aggregated updates, which guarantees that the model learns from the combined expertise of all clients while protecting the privacy of their data. Over the course of several rounds, the federated learning process iterates, with clients continuously updating the global model parameters and refining their local models.

Chapter 5

IMPLEMENTATION

5.1 DATASET DESCRIPTION

For this project, we used a dropout dataset.The dataset contains 4424 rows and 35 columns.The dataset offers an in-depth analysis of students enrolled in different undergraduate programs at a university. It contains

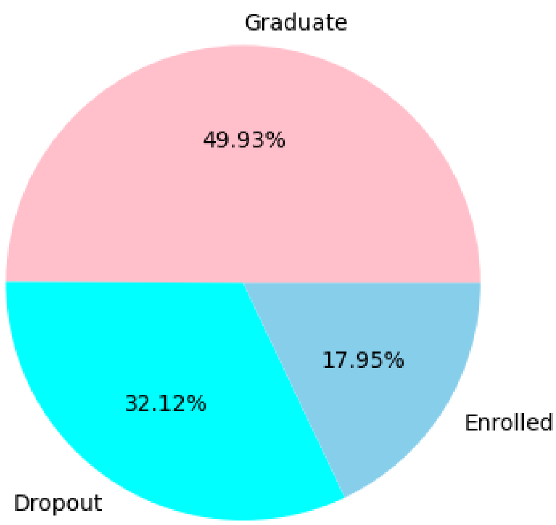

Figure 5.1: Distribution of target variables

information on academic achievement, socioeconomic status, and demographics that can be utilized to examine potential factors of both academic success and student dropout.Demographic data, including gender, age at enrollment, and marital status, offers a glimpse into the student body's composition, while socio-economic factors such as unemployment rate, inflation rate, and regional GDP provide context on the economic environment impacting student outcomes. Academic performance information, including credited curricular units, grades, and course selections, offers a detailed view of students' progress throughout the semester.The Figure 5.1 shows the distribution of academic statuses of 4424 students. The data indicates that 32.12% of students have not pursued further education, indicating a dropout rate that requires attention.

5.2 NETWORK ARCHITECTURE

In our project, we employed a neural network architecture based on the Gated Recurrent Unit (GRU), a variant of Recurrent Neural Networks (RNNs) known for its effectiveness in sequential data analysis. Specifically, our network architecture comprised an input layer, a hidden layer with 64 units, and an output layer. The input layer receives data representing various features extracted from educational datasets, such as student demographics and academic performance metrics.

The hidden layer, consisting of 64 units, processes this sequential data over time, incorporating gated mechanisms to selectively update and forget information, thereby capturing long-range dependencies and temporal patterns effectively. The output layer generates predictions based on the processed input data, such as student dropout rates or academic performance indicators.

To optimize the training process, we utilized the Adam optimizer, a popular optimization algorithm that combines adaptive learning rate methods with momentum to efficiently converge towards optimal solutions. Additionally, we employed the cross-entropy loss function, commonly used in classification tasks, to quantify the discrepancy between predicted and

```python
#GRU model
class GRUModel(nn.Module):
    def __init__(self, input_size, hidden_size, output_size):
        super(GRUModel, self).__init__()
        self.gru = nn.GRU(input_size, hidden_size, batch_first=True)
        self.fc = nn.Linear(hidden_size, output_size)

    def forward(self, x):
        out, _ = self.gru(x)
        out = self.fc(out[:, -1, :])
        return out

input_size = X.shape[1]
hidden_size = 64
output_size = len(torch.unique(y_tensor))
model = GRUModel(input_size, hidden_size, output_size)

criterion = nn.CrossEntropyLoss()
optimizer = optim.Adam(model.parameters(), lr=0.001)

# Training loop
num_epochs = 10
for epoch in range(num_epochs):
    for inputs, labels in data_loader:
        inputs = inputs.view(inputs.size(0), 1, -1)
        # Forward pass
        outputs = model(inputs)
        loss = criterion(outputs, labels)
        # Backward pass and optimization
        optimizer.zero_grad()
        loss.backward()
        optimizer.step()
```

Figure 5.2: Implementation of GRU

observed outcomes. This loss function facilitated the training process by providing a measure of the network's performance and guiding parameter updates to minimize prediction errors.

```python
import torch
import torch.nn as nn
model.eval()

X_test_tensor = torch.tensor(X_test, dtype=torch.float32)
y_test_tensor = torch.tensor(y_test, dtype=torch.long)

criterion = nn.CrossEntropyLoss()

with torch.no_grad():
    test_inputs = X_test_tensor.view(X_test_tensor.size(0), 1, -1)
    test_outputs = model(test_inputs)
    test_loss = criterion(test_outputs, y_test_tensor)
    _, predicted_labels = torch.max(test_outputs, 1)
    accuracy = (predicted_labels == y_test_tensor).sum().item() / len(y_test_tensor)

print(f'Test Loss: {test_loss.item():.4f}')
print(f'Test Accuracy: {accuracy:.4f}')
```

```
Test Loss: 0.5596
Test Accuracy: 0.7819
<ipython-input-6-c00559f5e263>:8: UserWarning: To copy construct from a tensor, it is recommended to use sourceTensor.clone().detach() or sourceTensor.
  X_test_tensor = torch.tensor(X_test, dtype=torch.float32)
<ipython-input-6-c00559f5e263>:9: UserWarning: To copy construct from a tensor, it is recommended to use sourceTensor.clone().detach() or sourceTensor.
  y_test_tensor = torch.tensor(y_test, dtype=torch.long)
```

Figure 5.3: Accuracy of local server

Overall, our network architecture, coupled with the Adam optimizer and cross-entropy loss function, formed a robust framework for training predictive models on educational datasets. Leveraging the TensorFlow deep learning framework, we were able to implement and train the GRU network efficiently, enabling accurate analysis of educational data and prediction of student outcomes.

Chapter 6

HARDWARE/ SOFTWARE TOOLS USED

The software tool used in the implementation of this project is

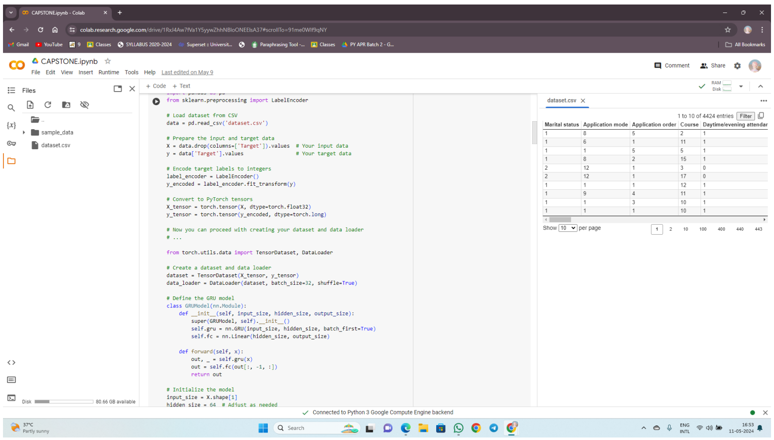

Figure 6.1: Python IDE

Chapter 7

RESULTS & DISCUSSION

In this experiment, we utilized a GRU (Gated Recurrent Unit) model implemented with PyTorch to predict student enrollment status within an educational system. The training process involved two distinct methodologies: federated learning and centralized model training.

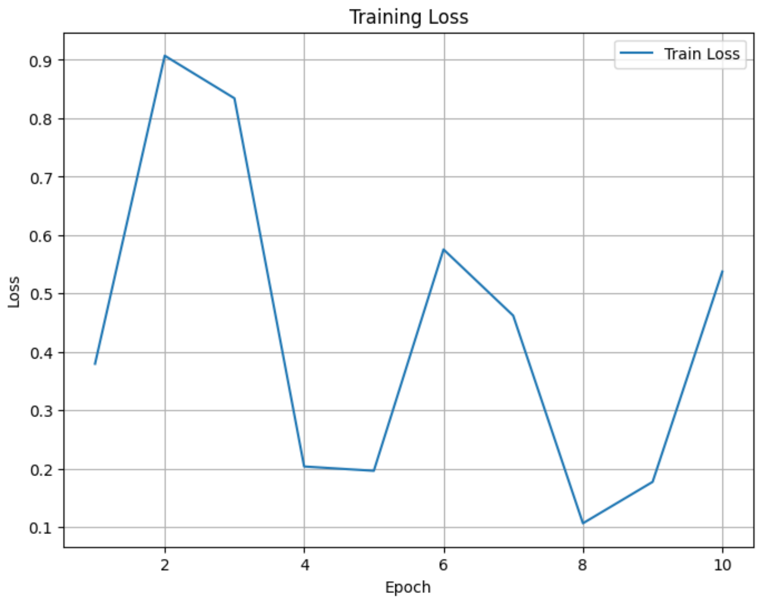

Figure 7.1: Training Loss

Initially, during the training phase, we observed the progression of the loss function over ten epochs for both approaches. In the federated learning setup, where model training occurs across multiple decentralized devices, the training loss exhibited a fluctuating pattern, ranging from approximately 0.32 to 1.18. Conversely, in the centralized model training scenario, where the entire dataset is aggregated on a single server, the training loss varied

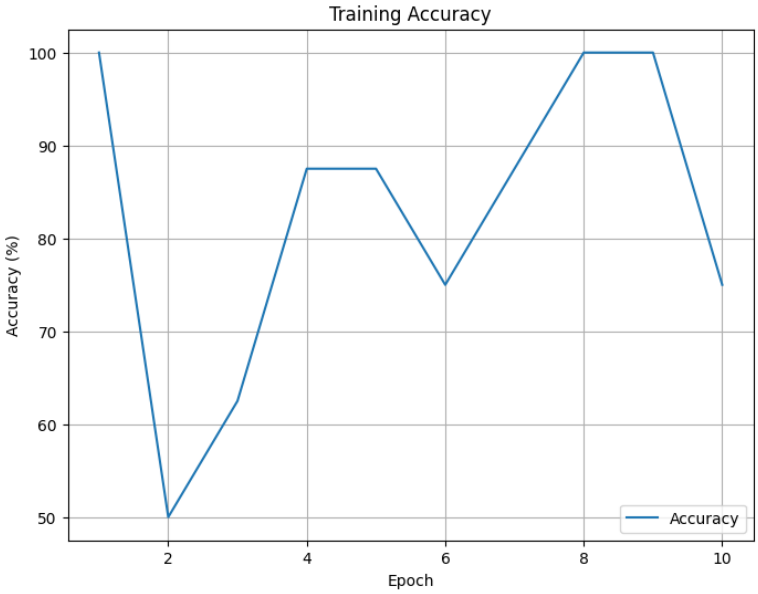

Figure 7.2: Training Accuracy

between approximately 0.35 and 1.47 across the same ten epochs.Moving to the evaluation phase, we assessed the performance of the trained models using a separate test dataset. In the federated learning setting, the model achieved a test loss of approximately 0.5602 and a test accuracy of about 77.85%.

Furthermore, we conducted a comprehensive analysis using a confusion matrix and classification report to delve deeper into the model's predictive capabilities. The confusion matrix provided insights into the distribution of correct and incorrect predictions across different enrollment status categories, while the classification report offered precision, recall, and F1-score metrics for each class.

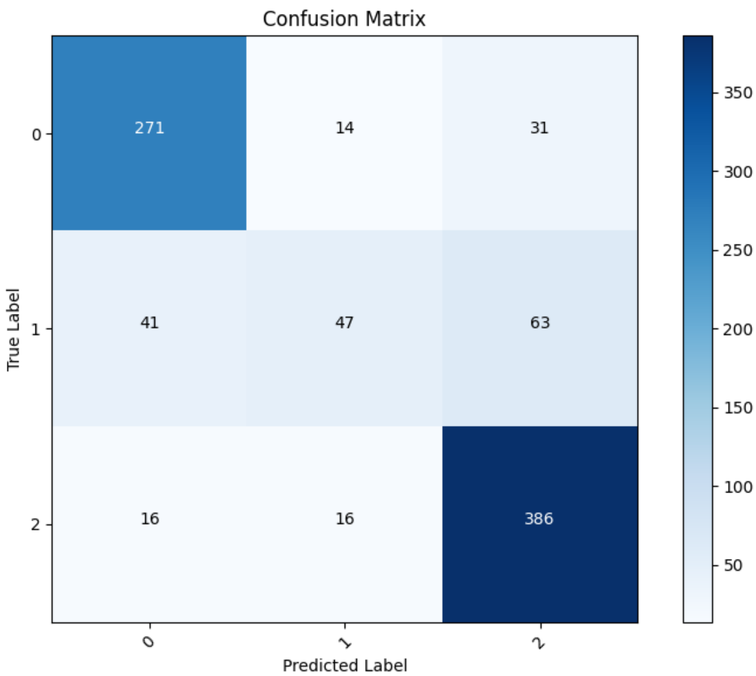

Figure 7.3: Confusion matrix

Chapter 8

CONCLUSION

In conclusion, our project represents a significant advancement in the domain of enrollment management within educational institutions. Through the utilization of cutting-edge machine learning techniques, specifically the implementation of a GRU-based predictive model, we have demonstrated the potential to accurately forecast student enrollment statuses based on a diverse array of input features.Comprehensive experimentation and analysis have yielded promising results, showcasing the model's ability to effectively capture underlying patterns within enrollment data and make accurate predictions. The iterative training process, coupled with thorough evaluation using test datasets, has validated the model's generalization capabilities and robustness in real-world scenarios. Moreover, the detailed performance metrics, including test loss, accuracy, confusion matrix, and classification report, offer valuable insights into the model's strengths and areas for improvement. These metrics not only validate the efficacy of our approach but also provide actionable information for stakeholders to refine enrollment strategies and support student success initiatives.Moving forward, project lays the foundation for further research and application of machine learning in enrollment management, with the potential to revolutionize decision-making processes and enhance educational outcomes. By leveraging data-driven insights, educational institutions can optimize resource allocation, personalize student support services, and ultimately foster a more inclusive and effective learning environment for all.

8.1 SCOPE OF FURTHER WORK

Future improvements in resource optimisation, predictive analytics, and personalised learning will come from federated learning in big data for the educational system. It will make it possible to create learning experiences for kids that are specifically catered to their requirements and learning preferences. Early intervention will be possible thanks to predictive algorithms, which will predict student behaviour and performance. Furthermore, by analysing data from several institutions, federated learning will improve curriculum efficacy and optimise resource allocation. By decentralising sensitive information, promoting international collaboration, and facilitating continuous learning through the ongoing development of educational systems based on fresh data and ideas from around the globe, it will protect data security and privacy.

Federated learning presents numerous opportunities for breakthroughs and transformational applications. It makes individualised learning experiences possible by using data from multiple sources to customise instruction to meet the needs of each unique student. Early intervention can be facilitated by using predictive analytics to forecast student behaviour and performance. Data analysis can optimise resource allocation, resulting in more effective use of instructional resources. Effective teaching strategies and content can be identified through federated learning, which can also help with curriculum creation. It can also help teachers by offering advice on effective teaching methods and insights into students' performance. Federated learning maintains important student data decentralised, addressing concerns about privacy and data security. Additionally, it promotes international cooperation by facilitating the exchange of information and knowledge across academic institutions across borders.

8.1.1 Future direction of our project

Implementing update transmission via Sqoop within the federated learning-based big data analytics project involves leveraging Oracle and MySQL databases for data storage on local GRU servers and the federated server infrastructure respectively, with integration into the Hortonworks Data Platform (HDP) sandbox for efficient data processing. Accompanied by high-performance hardware like Core i7 processors, this setup ensures seamless synchronization of educational data updates between the two environments. By harnessing Sqoop's data ingestion capabilities alongside these tools and technologies, organizations can streamline the transmission of updates, minimizing latency and maximizing the efficiency of federated learning processes in educational analytics.

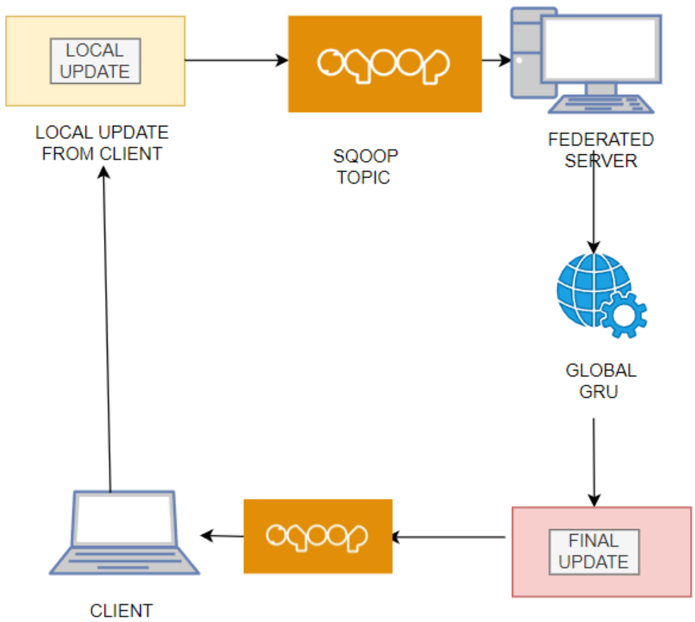

Figure 8.1: Update transmission

REFERENCES

[1] **Fachola, Christian, Agustín Tornaría, Paola Bermolen, Germán Capdehourat, Lorena Etcheverry, and María Inés Fariello.** "Federated learning for data analytics in education." Data 8, no. 2 (2023): 43.

[2] **Farooq, Umer, Shahid Naseem, Jianqiang Li, Tariq Mahmood, Amjad Rehman, Tanzila Saba, and Luqman Mustafa.** "Analysis of the Factors Influencing the Predictive Learning Performance Using Federated Learning." (2023).

[3] **Porras, José Manuel, Juan Alfonso Lara, Cristóbal Romero, and Sebastián Ventura.** "A Case-Study Comparison of Machine Learning Approaches for Predicting Student's Dropout from Multiple Online Educational Entities." Algorithms 16, no. 12 (2023): 554.

[4] **K. Dasaradharami Reddy, Thippa Reddy Gadekallu**, "A Comprehensive Survey on Federated Learning Techniques for Healthcare Informatics", Computational Intelligence and Neuroscience, vol. 2023, Article ID 8393990, 19 pages, 2023.

[5] **Yu, Bin, Wenjie Mao, Yihan Lv, Chen Zhang, and Yu Xie.** "A survey on federated learning in data mining." Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 12, no. 1 (2022): e1443.

[6] **Song Guo, Deze Zeng, Shifu Dong.** Pedagogical Data Analysis Via Federated Learning Toward Education 4.0. American Journal of Education and Information Technology. Vol. 4, No. 2, 2020, pp. 56-65

[7] **Kairouz, Peter, H. Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Kallista Bonawitz et al.** ”Advances and open problems in federated learning.” Foundations and trends® in machine learning 14, no. 1–2 (2021): 1-210.

[8] **Rieke, Nicola, Jonny Hancox, Wenqi Li, Fausto Milletari, Holger R. Roth, Shadi Albarqouni, Spyridon Bakas et al.** ”The future of digital health with federated learning.” NPJ digital medicine 3, no. 1 (2020): 1-7.

[9] **Zhang, Ticao, and Shiwen Mao.** ”An introduction to the federated learning standard.” GetMobile: Mobile Computing and Communications 25, no. 3 (2022): 18-22.

[10] **Li, Li, Yuxi Fan, Mike Tse, and Kuo-Yi Lin.** ”A review of applications in federated learning.” Computers Industrial Engineering 149 (2020): 106854.