INTERNATIONAL JOURNAL OF SOCIAL SCIENCES BULLETIN
ISSN: (E) 3007-1917 (P) 3007-1909
Volume 4, Issue 2, 2026

HUMAN CAPITAL AND WELL-BEING METER BASED ON FEDERATED LEARNING POWERED SENTIMENT ANALYSIS

Muhammad Shaheer[*1], Abroo Nazar[2], Tayyaba Arooj[1], Muhammad Hassan Farid[1], Muzamil Malik[1], Muhammad Inam-Ur-Rehman[1], Talha Riaz[1], Syed Asad Abbas[1], Sania Ishaq[1]

[1]*Hamdard University Islamabad Campus*
[2]*AL HAMD Islamic University*

Corresponding Author: *
Muhammad Shaheer
DOI: https://doi.org/10.5281/zenodo.18537536

Received	Accepted	Published
07 December 2025	22 January 2026	06 February 2026

ABSTRACT

The true well-being of a person does not only mean money, but it is the overall quality of life, the social, emotional, and psychological state of a person. All these emotions are usually reflected in common speech. Though a mere textual analysis will inform us of the emotional tone by sentiment analysis, a whole picture of well- being means an appreciation of the context in the emotion. This is offered in our well-being meter, which incorporates a capital prediction component that links the feeling to the area of life capital such as financial or social, that it was caused by. Such a general approach is necessary to have a comprehensive perspective. Nevertheless, the main risk of privacy is the old-fashioned and centralized means of analyzing this form of sensitive data. This paper proposes FedBiLSTM-Net, a secure approach based on Federated Learning (FL) and a Bidirectional Long Short-Term Memory (Bi-LSTM) network. This strategy stores the raw and sensitive text in local machines, and the clients share model updates with a central server. This decentralized data ensures privacy and integrity of data. We constructed a dataset of 336,029 text records based on both the public and clinical datasets such as CLPsych Shared Task and DAIC-WOZ datasets, to include sentiment and other types of capital. We adopted Bi-LSTM into our FL model after ensuring that it outperformed in the first test. We trained the model on 20 different independent data groups (non-IID clients) in 10 rounds using the Flower framework and FedAvg. It considers sentiment and capital prediction simultaneously, which makes it less complicated. The last model had a centralized accuracy of 98.59%. These findings indicate that a privacy-aware model can be trained using Federated Learning to effectively learn intricate well-being patterns using a diversified, distributed dataset. Our model FedBiLSTM-Net provides a viable privacy-first well-being existence tracker.

Keywords:

1 Introduction

A well-being meter is a mechanism that evaluates human well-being based on analyzing textual data, assessing the emotional and psychological conditions. In recent years, digital platforms generate vast amounts of textual data. This user-generated content includes emotional and psychological cues as well [1] which reflect mental well-being of a person. Human mental well-being is widely recognized as an important indicator of quality of life [2]. Therefore, it is crucial to develop reliable well-being metrics. However, traditional approaches

which include surveys, clinical assessments or wearable sensors have some limitations. For instance, clinical assessments or surveys are considered infrequent and costly [2]. Similarly, self-report questionnaires are unreliable because they can be biased [3]. These limitations underscore the need for novel well-being measurement methods that are scalable and effective. User generated data is a valuable resource that can overcome the limitations of traditional methods. Additionally, sentiment analysis on such data enables assessment of well-being in a practical manner without any additional efforts from individuals. However, analyzing sensitive user-generated data with centralized approaches raises privacy concerns.

Earlier studies on the sentiment analysis for well-being relied on shallow machine learning methodologies such as Na¨ıve Bayes (NB), Logistic Regression (LR) and Support Vector Machines (SVM). The systematic reviews reported F1-scores often below 60% for these approaches in health-related applications [4]. Deep learning models such as Convolutional Neural Networks (CNNs) and Long Short Term Memory networks (LSTMs) improved performance by modeling semantic and sequential dependencies [5]. Bi-LSTMs can effectively capture sequential and contextual dependencies in textual data. However, these methods were deployed in centralized settings where raw user data had to be pooled into a single repository. This raised serious ethical and legal concerns when applied to sensitive human well-being data [6]. Decentralized techniques such as Federated learning offers a privacy-preserving solution [7]. Federated learning allows for local training on user devices and only model updates are sent to the central server [8]. These updates are aggregated by the central device to improve the performance of the global model. Afterwards, these aggregated updates are shared back with the devices. Recent studies demonstrated the use of FL is promising in mental state detection [9] and emotion recog- nition [10]. Despite this progress, most of the prior studies either used lightweight models with limited contextual understanding or computationally heavy transformers.

Moreover, there are several challenges associated with implementing FL for well-being monitoring. For instance, data across users is usually non-independent and non-identically distributed (non-IID) [11]. In addition to this, the variations due to demography and culture present in the well-being data makes the model generalization more difficult [12]. During the model training, sensitive information can leak through the gradient. This risk can be mitigated by using methods like differential privacy that add noise. However, it can result in reduced model accuracy [13]. Similarly, cryptographic methods like secure multiparty computation and homomorphic encryption are used to enhance confidentiality, however, they need heavy computation that mobile and wearable devices often cannot handle effectively [14].

The complexity of natural language makes it more challenging. Although publicly available datasets like CLPsych Shared Task and DAIC-WOZ capture a wide range of natural language expressions related to emotional and psychological states, it includes natural language intricacies i.e., negation, sarcasm, metaphors and emoji use. This can complicate sentiment analysis and mislead various machine learning models [15] [16]. Recent studies demonstrated that emoji embeddings integrated into deep language models like Bidi- rectional Encoder Representations from Transformers (BERT) can significantly improve sentiment classification performance as they enable it to capture multimodal cues [16].

Moreover, accessing data presents additional challenges in human well-being research. Common sources which include WHO-5 Well-being Index and Patient Health Question- naire (PHQ-9) and behavioral data from smartphones and digital wearables [17] are strictly protected by privacy regulations like General Data Protection Regulation (GDPR) and the Health Insurance Portability

and Accountability Act (HIPAA). These restrictions have consequences such as limitation in centralized data sharing and barriers to large scale analysis [18].

To address these gaps, this study proposes **FedBiLSTM-Net**, a novel privacy-preserving and context aware well-being meter that integrates Bi-LSTM with FL. Unlike prior work, our framework explicitly evaluates robustness under 20 non-IID clients. It also incorpo- rates strategies for handling challenging linguistic cues and compares Bi-LSTM with tradi- tional baselines like LR and SVM. The proposed model demonstrates robust performance in both centralized and federated settings. By combining Bi-LSTM's ability to capture sequential and contextual dependencies in text with FL's decentralized training, this work contributes an efficient, scalable, and privacy-preserving solution for reliable well-being monitoring without requiring resource-heavy transformer architectures. Additionally, the proposed novel well-being meter can predict both sentiment and capital. Sentiment anal- ysis captures the emotional polarity i.e., positive or negative but does not explain why a person feels that way. The capital such as economic, social, emotional and psycholog- ical resources provides the contextual domain that influences an individual's well-being. So, by predicting both sentiment and capital, our model links emotional expression with underlying context. This multi prediction approach provides a more comprehensive and meaningful understanding of human well-being as compared to conventional sentiment models.

2 Literature Review

2.1 Traditional Machine Learning Approaches

Sentiment analysis, a core task in natural language processing (NLP), aims to classify emotions from textual data. Early approaches relied heavily on traditional machine learning models such as Support Vector Machines (SVM) [1], Logistic Regression (LR) [2] and Na¨ıve Bayes. These models were expert in handling structured and high-dimensional text data, but it requires careful feature engineering, limiting their capability to capture subtle emotional or psychological indications and signs in natural language.

SVMs are majorly effective for binary text classification performing well on high dimensional and sparse datasets such as the data of social media [3]. Some feature selection methods like Chi-Square and n-gram models have improved SVM performance [6], while kernel functions allow SVMs to separate classes in higher dimensions without overfitting [33]. For example, Povoda et al. introduced a language-independent SVM model for multilingual sentiment analysis, achieving over 88% accuracy across multiple languages using distributed learning on 112 processors [34].

Logistic Regression (LR) has also been applied successfully to multi-class sentiment clas- sification. Pranckevičius and Marcinkevičius demonstrated that adding part-of-speech features and increasing the size of training data improved LR accuracy [7]. Indra et al. developed a web-based application using LR to classify tweets into topics like health, mu- sic, sport, and technology, achieving 92% accuracy with 1800 labeled tweets per topic [8]. Despite their effectiveness, traditional ML models face challenges with large-scale, het- erogeneous and sensitive data. Centralized storage requirements pose privacy issues, and these models struggle with scalability and flexibility in dynamic real-world environments.

Table 1: Comparing Sentiment Analysis with SVM, LR, and FL

Aspect	Support Vector Machine (SVM)	Logistic Regression (LR)	Federated Learning (FL)
Data Handling	Requires centralized dataset	Requires centralized dataset	Can work on decentralized/local data
Scalability	Efficient with small/medium data	Moderate	High (Fit for large-scale, cross-device training)
Privacy	Low (sharing of data is required)	Low (sharing of data is required)	High (data remains on devices)
Flexibility	Restricted for dynamic/heterogeneous data	Restricted feature expansion	Robust, adapts to diverse, distributed environments

1.1 Federated Learning Foundations

Privacy concerns in centralized machine learning have motivated the adoption of Federated Learning (FL). Introduced by Google [2], FL enables collaborative model training without moving raw data from local devices. Instead, a central server distributes an initial global model to clients, who train it locally and send only updated parameters such as (weights and gradients) back. The server aggregates these updates, typically via weighted averaging to refine the global model iteratively. This approach preserves user privacy while leveraging diverse, distributed datasets.

2.1.1 Federated Learning Algorithms

Several algorithms enable decentralized training in federated environments:

1. **Federated Averaging (FedAvg)** is particularly well-suited for cross-device scenarios involving mobile devices and IoT systems [1]. FedAvg aggregates local model updates from distributed clients to build a global model while minimizing communication costs. Each client trains locally, then the central server computes a weighted average to update.

$$w^{t+1} = \sum_{k=1}^{K} \frac{n_k}{n} w_k^{t+1} \qquad (1)$$

where: w^{t+1} is the updated global model at communication round $t + 1$. K represents the total number of clients, n_k is the number of samples on client k and n is the total number of samples across all clients ($n = \sum_{k=1}^{K} n_k$) and w_k^t represents the local model parameters from client k at round t.

2. **FedProx** extends FedAvg to handle data heterogeneity more effectively [1]. Unlike FedAvg, FedProx allows variable local update iterations and adds a proximal term to the optimization objective. This term prevents excessive parameter drift when clients have significantly different data distributions.

$$\min_{w} h F_k(w) + \frac{\mu}{2}\|w - w^t\|^{2^i} \qquad (2)$$

where: $F_k(w)$ is the local loss function for client k and w denotes the local param- eters being optimized also w^t is the current global model where $\mu \geq 0$ controls the regularization strength and $|w - w^t|^2$ measures the L_2 distance between local and global parameters.

3. **Decentralized Stochastic Gradient Descent (DSGD)** adapts traditional SGD for federated settings. In DSGD, clients update models using local data and pe- riodically exchange gradients with neighbors rather than a central server. This architecture suits large-scale FL systems with limited bandwidth [9].

4. **FedSGD vs. FedAvg**: Research comparing FedAvg and FedSGD [11] reveals important performance trade-offs in semi-asynchronous federated learning. While FedSGD achieves better accuracy and faster convergence, it exhibits 58% more oscillation than FedAvg, occasionally causing numerical instability (NaN values) during training.

5. **q-FedAvg** addresses fairness by modifying the standard FedAvg objective to re- duce performance disparities across heterogeneous clients [1]. It penalizes poorly performing clients to balance overall model performance, though this may create trade-offs between fairness, privacy, and accuracy [10].

6. **Per-FedAvg (Personalized FedAvg)** focuses on client-specific personalization through local fine-tuning, tailoring models to individual client needs rather than maintaining a single global model [1].

1.2 Deep Learning for Sentiment Analysis

Deep learning architectures integrate effectively with FL, improving sentiment analysis while preserving privacy. CNN-based FL models have achieved up to 88% accuracy across multiple emotional categories without centralizing raw data [12]. Other studies have explored GRU, CNN, and Bi-LSTM in federated environments, showing that GRU and
Bi-LSTM often outperform CNN in accuracy, while LSTM models remain computationally efficient for sequential data tasks [25], [28].
Enhanced architectures like EnLNet have been proposed for multimodal sentiment anal- ysis, maintaining high precision (> 95%) while learning from heterogeneous sources [13]. FL-enabled LSTM models have shown strong performance in tasks like spam detection (99.55% accuracy) and sarcasm detection, demonstrating their suitability for sequential and context-dependent text [27], [29].
Federated LSTM applied to spam SMS detection on the Multilingual Spam Data dataset achieved 99.55% accuracy, 98.24% recall, and 98.25% F1-score [27]. These results demon- strate LSTM effectiveness for sequential text classification. Federated LSTM has also succeeded in complex tasks like sarcasm detection [28]. While BiLSTM achieves superior performance, standard LSTM effectively captures long-term dependencies through gating mechanisms. LSTM's ability to model both short-term and long-term temporal patterns makes it

particularly suitable for sentiment analysis [29].

1.3 FL Applications in Sentiment Analysis

Federated learning performs well with Independent and Identically Distributed (IID) data but degrades under non-IID conditions [15]. Analysis of FFNNs, LSTMs, and transform- ers on the Sentiment140 dataset (1.6 million labeled tweets) revealed that label imbalance hurts accuracy more than data quantity imbalance. Transformers showed greater robust- ness to non-IID scenarios compared to FFNNs and LSTMs.

Work on sentiment analysis across social media platforms explored various FL paradigms, including horizontal and vertical federated learning, as well as frameworks like TensorFlow Lite, PySyft, NVIDIA FLARE, FedML, Flower, and FLSys [16]. The FL-based system achieved 94.86% accuracy while preserving privacy.

A systematic review of 36 federated learning studies [17] examined FL from multiple per- spectives: model architecture, network infrastructure, and data privacy and security. The review confirmed FL improves both accuracy and efficiency in sentiment analysis. How- ever, it identified a gap in comprehensive evaluation of privacy and security implications for federated sentiment analysis systems.

OpenFedLLM represents recent developments in training Large Language Models using federated learning [19]. This framework combines multiple FL algorithms with federated instruction tuning and value alignment, incorporating over 30 evaluation metrics across 8 domain-specific datasets. FL-trained models outperformed locally trained LLMs, with federated fine-tuned Llama2-7B surpassing GPT-4 on financial benchmarks.

1.4 FL Challenges

Federated learning introduces unique challenges:

- **Privacy**: Although FL keeps raw data local, risks remain due to potential leakage via gradients or updates.
- **Non-IID Data**: Performance often degrades when clients have different data dis- tributions, as seen with label imbalance on Sentiment140 [15].
- **Communication Efficiency**: Algorithms like FedAvg and DSGD attempt to re- duce communication overhead, but trade-offs exist.
- **Fairness**: Addressed by algorithms like q-FedAvg, but it may reduce accuracy or privacy guarantees [10].
- **Scalability and Infrastructure**: As seen in the review [17], system-level chal- lenges include bandwidth, synchronization, and deployment frameworks.

2 Methodology

This section outlines the process of the creation of a well-being meter that can be used to assess the capital and sentiment of an individual. This method starts with the systematic data collection and is preprocessed to maintain the data quality and consistency followed by the Exploratory Data Analysis (EDA) to determine the patterns, trends, and distri- butions of the data. Appropriate metrics are used to validate effectiveness of a model's performance. Lastly, a simulation of federated learning (FL) is performed to illustrate the training of a model through decentralization and maintain privacy of the data.

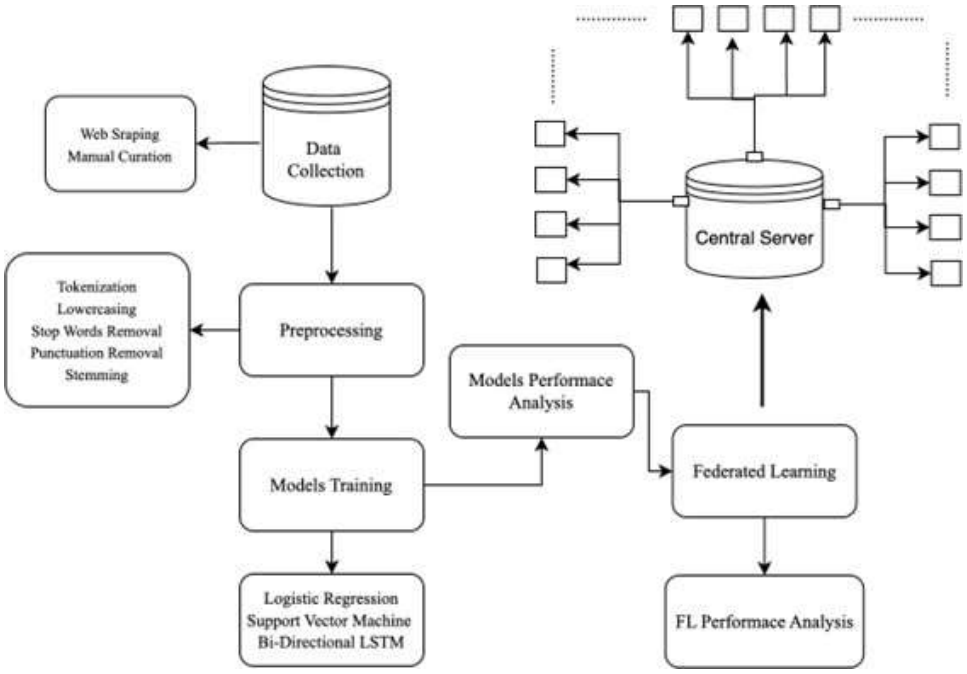

Figure 1: Proposed Methodology

2.1 Data Collection

Data had been gathered through web scraping and manual augmentation. Data were extracted from dynamic organizational web pages and finally, the process of manual cu- ration balanced the coverage of sentiments and capital types. The ultimate corpus was 336,029 systematic records of individual capital and sentiment.

It has a combination of text-based attributes, numeric (e.g., age, user IDs), categorical (e.g., day of week, sentiment), temporal (e.g., date, time), and geographic (e.g., country) variables shown in Table 2. The aim of the EDA will be to examine the structure of the dataset, evaluate the quality of data, and mention preliminary trends before utilizing advanced analytical models.

Table 2: Types of data

Column Name	Data Type	Description
Capitals Text Generation	Text	Raw text data related to capital generation
Sentiment Label	Categorical	Sentiment classification (positive or negative)
Predicted Capital	Text	Predicted capital names
User-Id	Numeric	Unique identifier for each user
Social-Accounts	Text	Information about users' social media accounts
Date	Date	Date of data recording
Time	Time	Time of data recording
Day	Categorical	Day of the week when the data was

		recorded
Country	Text	Name of the country related to the capital data
Age-Of-User	Numeric	The age of the user providing the data

Capitals in this research are the multi-dimensional assets that define the welfare of an individual or society- economic, social, intellectual and spiritual. Sentiment analysis does not represent the drivers of well-being but only the emotional tones. The well-being meter proposes a comprehensive view of well-being (both personal and social) by modeling sentiment (how people feel) and capital distribution (what aspects of life they emphasize). Federated Learning (FL) makes it possible to accomplish such an analysis in a privacy- preserving, decentralized fashion so that sensitive well-being data can be kept locally and yet contribute to global knowledge.

We consider 12 forms of capital:

1. Economic Capital- Money and financial resources.
2. Social Capital - Networks and Relationships.
3. Emotional Capital - The ability to be emotionally aware and to manage them.
4. Digital capital Capability and access to digital technology and resources.
5. Cultural Capital - Knowledge and values and habits that are acceptable in society.
6. Creative Capital - Power to become creative and develop new ideas.
7. Temporal Capital - Time to spend at work, build or maintain relationships.
8. Physical Capital - It is a physical asset that includes equipment, tools or property.
9. Intellectual Capital- Value adding knowledge and ideas.
10. Organizational Capital - Processes and systems which render organizations viable.
11. Experiential Capital- Experiential knowledge and wisdom.
12. Human Leadership Capital - The human aspect in terms of skills and leadership that drives others to success.
13. Religious capital - Community and valuing religiousness.
14. Psychological Capital - Good traits like hope, confidence and resilience.
15. Reputational Capital- Credibility and respect of other people.
16. Spiritual Capital- moral strength and inner motivation.

Text segments are classified into these categories through supervised annotation. Each sentence in the corpus is labeled according to its capital type.

2.2 Preprocessing

To eliminate unwanted data that influences model performance like biasness, variance, etc. the preprocessing of text involves removal of html tags, emojis, stop-word, normal- ization, tokenization and padding to make text sentences ready to analyze their capital and sentiments.

2.3 Exploratory Data Analysis (EDA)

This analysis will be important in the comprehension of the sentiments of people because it will not only point at their inner intention and behaviors but also provide useful points on the factors that shape their decision making hence help in providing an in-depth interpretation of the general results.

EDA has provided a consequential understanding of the data structure, distributions, and correlation amongst variables especially in relation to categories of sentiments, the type of predicted capital, and demographic traits. These results guaranteed

not only the clean and reliable data, but also the rich patterns, which could further feed the modeling and interpreting. In general, the exploratory stage provided a solid base to the further analysis as it revealed the tendencies, confirmed the integrity of the data, and preconditioned the research of how various types of capital are manifested and perceived in different social and cultural environments.

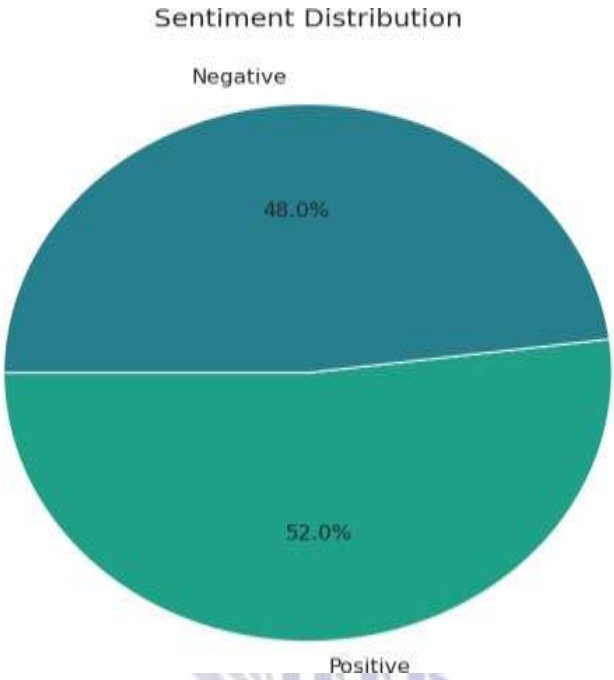

Figure 2: Sentiment distribution

The proportions of sentiment in the dataset are depicted in Figure 2. The positive sen- timents are 52.0% and the negative sentiments 48.0%. The small gap shows that there is a narrow band of opinions. This gives the two categories of sentiments sufficient to represent analysis of the dataset.

Figure 3: Most repeated words in positive and negative sentiments

In figure 3, comparison of negative and positive feelings, where negative ones focus on emotional stress, lack of tools and poor results, whereas the positive ones focus on time management, prior claims and purpose. This comparison emphasizes that negative views are being formed in the light of the challenge and inefficiency, as opposed to the posi- tive

views, which are formed in the light of organization, productivity, and meaningful interaction.

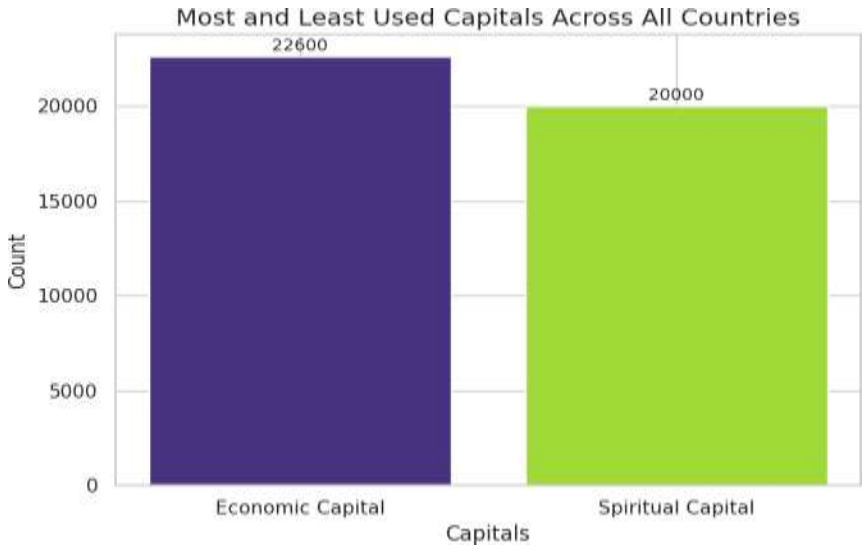

Figure 4: Count of Most and Least Used Capitals Across All Countries

The figure 4, shows the most and least used capitals among different countries as economic capital (a count of 22,600) and spiritual capital (a count of 20,000) respectively are most and least stressed. It means that the world has a stronger tendency to be oriented on economic and not spiritual aspects implying that in the sentiment analysis data the dialogues and statements are more focused on the material and financial but not on the spiritual.

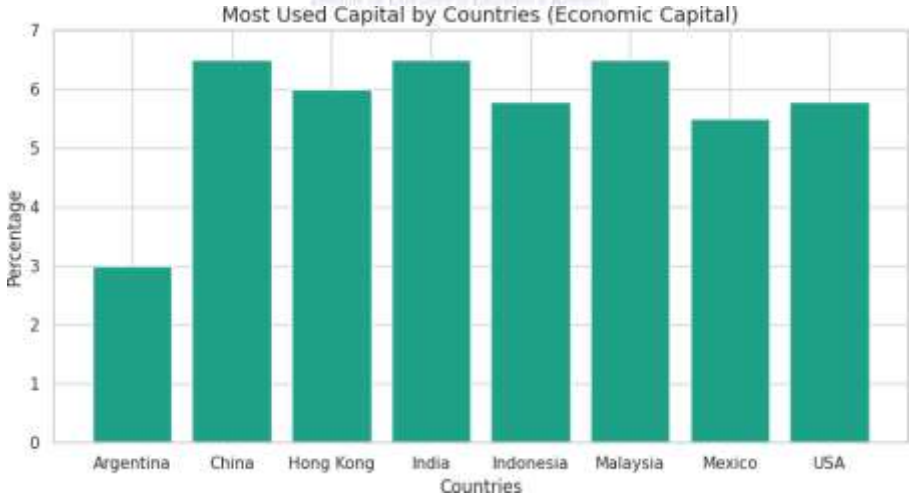

Figure 5: Percentage of Most Used capital (Economic Capital) by each country

The figure 5 indicates the percentage utilization of the economic capital in eight regions (Argentina (~ 3%), China (~ 6%), Hong Kong (~ 5.5%), India (~ 6%), Indonesia (~ 5.5%), Malaysia (~ 6%), Mexico (~ 5%) and USA (~ 4.5%)) where China, India and Malaysia top the list with about 6 percent usage. This is probably the ratio of eco- nomic capital (e.g., financial resources or investments) that is actively used,

but the exact definition is vague without additional context and may imply investment trends or the efficiency of capital. The detachment of Hong Kong and China would imply specialization in major economic centers, indicating disparities in capital utilization that may be due to variations in economic policy or level of development by October 05, 2025.

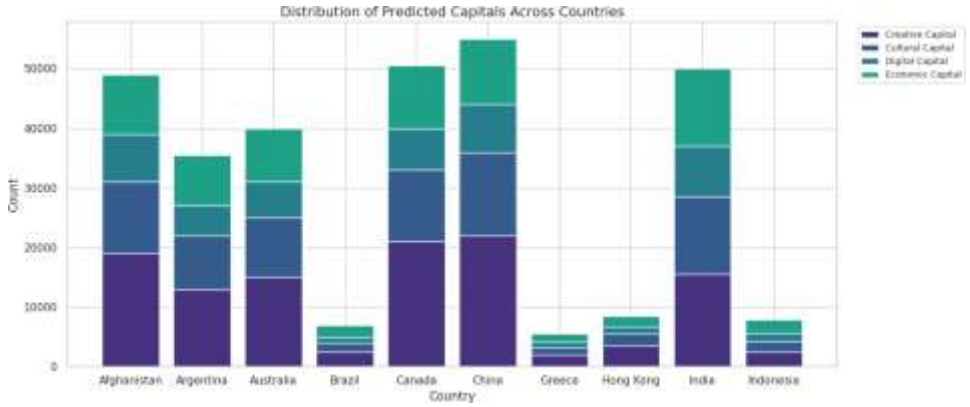

Figure 6: Highest and Lowest Used Capitals by Top 10 Countries

The figure 6, shows significant heterogeneity in the predicted capital composition of these nations, Greece seems dominated by Creative and Digital Capital, whereas countries like Afghanistan and Brazil seem to have a more dispersed distribution of their capital types.

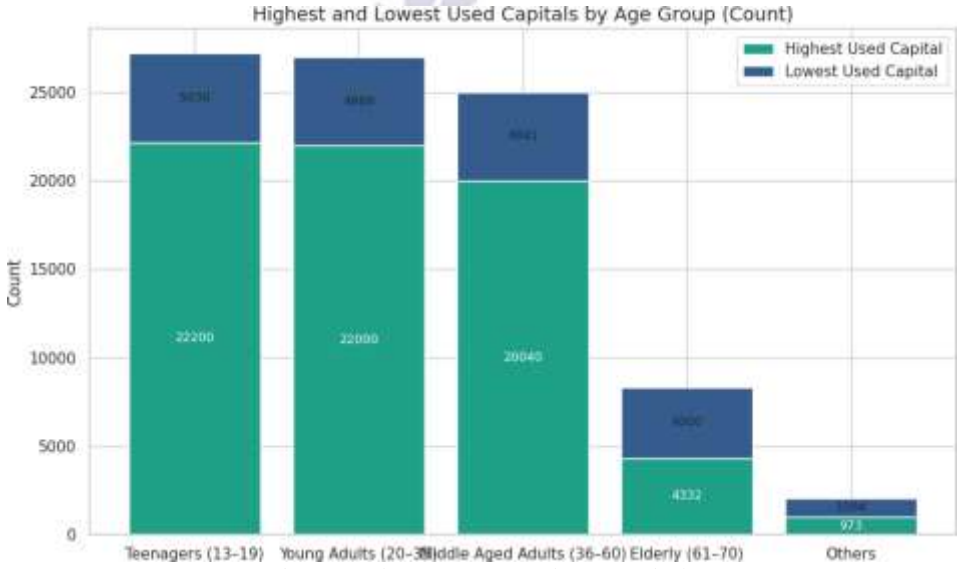

Figure 7: Highest and Lowest Used Capitals by Age Group

In figure 7, teenagers (ages 13 to 19), emotional capital (26.1%) is the most used resource, whereas reputational capital (5.9%) is the least used; Economic capital is more important to young adults (20–35) than human leadership capital (4.7%); middle-aged adults (36–50) rely heavily on spiritual capital (24.4%), with digital capital (6.0%) be- ing the least used; the elderly (51–70) favor reputational capital (8.0%) and temporal capital (8.9%). Indicating a generational shift from emotionally driven youth resources to economically and spiritually oriented adult capitals, and finally

to reputational and time-related concerns in later life.

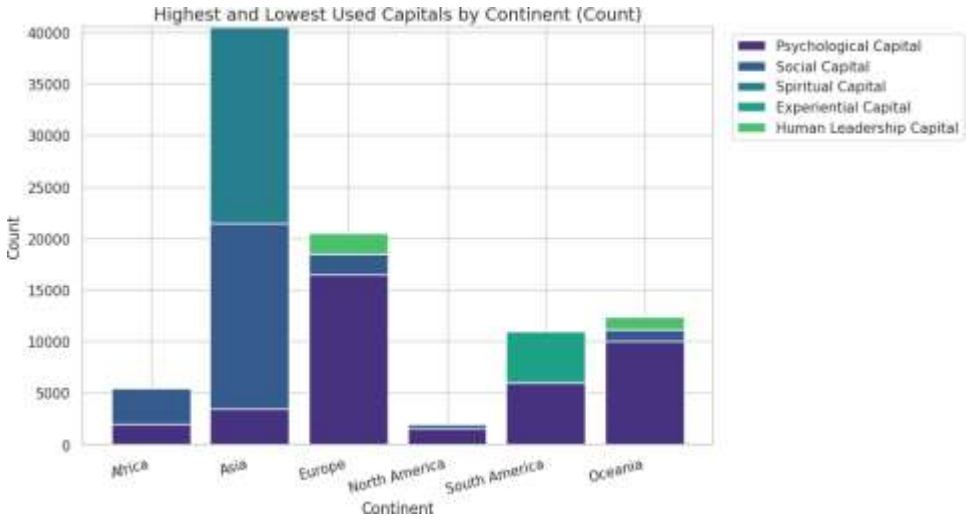

Figure 8: Highest and Lowest Used Capitals by Continents

The highest-used capitals on the continent are shown in Figure 8, with Asia leading with about 23,000 counts, primarily due to Spiritual Capital. Oceania (~ 10, 000) favors experiential capital, whereas Europe (~ 19, 000) places more importance on social and reputational capital. With the lowest score (~ 6, 000), Africa primarily relies on social and reputational capitals, indicating a global gradient from spiritual dominance in Asia to resources that are more experiential and leadership-focused elsewhere.

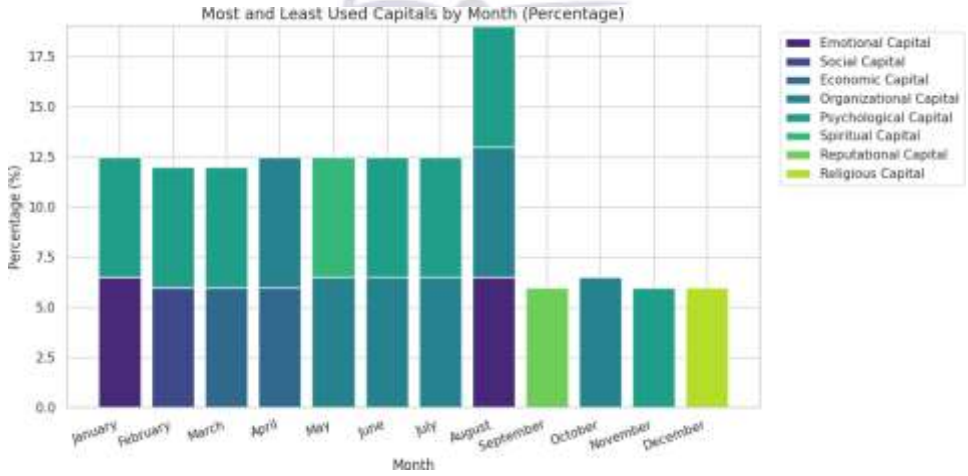

Figure 9: Most and Least Used Capitals by Month

Figure 9 visualizes the Percentage of the most and least used capitals over different months of the year, presenting information on how capital usage fluctuates throughout the year.

2.4 Models

For the Human Capital and Well-being meter, we were focused on getting the most reliable output. To ensure this, we proposed three (3) models so that we can compare their performance and choose the best one to carry forward with Federated Learning. The models that have been proposed are:

1. Logistic Regression (LR)

2. Support Vector Machine (SVM)
3. Bidirectional Long Short-Term Memory (Bi-LSTM)

We used the same dataset for all three models including proposed and trained them under similar evaluation criteria. All three models performed well, and they were fine-tuned to increase their performance. The comparison of these three (3) models are provided in this section.

1. **Logistic Regression (LR):** The baseline linear classifier was implemented using scikit-learn's LogisticRegression module with L2 regularization. Text features were extracted using Term Frequency-Inverse Document Frequency (TF-IDF) vectoriza- tion with a maximum feature limit of 5,000 terms and bi-gram support. The model utilized maximum likelihood estimation for parameter optimization with conver- gence tolerance set to 1e-4 and maximum iterations of 1,000.

2. **Support Vector Machine (SVM):** A non-linear SVM classifier was deployed using the Radial Basis Function (RBF) kernel for complex pattern recognition in high-dimensional feature space. The model operated on identical TF-IDF represen- tations as the logistic regression, maintaining consistent feature engineering across traditional approaches. Probability estimates were enabled to facilitate comprehen- sive performance comparison including precision-recall analysis.

3. **Bidirectional Long Short Term Memory (Bi-LSTM):** Bidirectional LSTM (Bi-LSTM) model was developed to do sentiment and capital classification by using shared representations to improve the tasks. The architecture consisted of a 64-dimensional embedding space with spatial dropout (0.3) which was then followed by a bidirectional 64 units LSTM with another dropout that minimized overfitting. Sentiment and capital prediction were done using two parallel softmax outputs, which allowed multi-task learning. The model was trained using the Adam optimizer and categorical cross-entropy loss in 10 epochs with a batch size of 32 and did not exhibit any major sensitivity issues to contextual dependencies or subtle semantics, so it should be suitable when dealing with less context-sensitive and more complex text classification problems.

2.4.1 Model Performance Analysis and Visualization

Table 3 shows how the models perform on both Sentiment and Capital classification prob- lems in comparing Bi-LSTM, Logistic Regression, and SVM on four standard evaluation measures of accuracy, precision, recall, and F1-score.

Table 3: Model Comparison

Model	Task	Accuracy	Precision	Recall	F1-Score
Bi-LSTM	Sentiment	99.41%	99.41%	99.41%	99.41%
Bi-LSTM	Capital	99.75%	99.98%	99.97%	99.97%
Logistic Regression	Sentiment	96.85%	96.89%	96.86%	96.87%
Logistic Regression	Capital	98.65%	99.88%	99.89%	99.88%
SVM	Sentiment	97.86%	97.86%	97.85%	97.86%
SVM	Capital	98.89%	99.93%	99.96%	99.94%

Bi-LSTM model had an accuracy of 99.41% in sentiment classification tasks and 99.75% in capital classification tasks as shown in Table 3. In the case of sentiment analysis, precision, recall, and F1-score is 99.41%. Precision in the capital classification task stood at 99.98%,

recall is 99.97%, and F1-score 99.97%. These measures indicate that the Bi-LSTM has almost perfect classification accuracy and multi-class classification performance in both binary and multi-class classification.

Logistic Regression achieved an accuracy on sentiment classification and capital classification tasks were 96.85% and 98.65% as shown in Table 3. The precision, recall and F1-score are 96.89%, 96.86% and 96.87% respectively in the sentiment analysis. In terms of capital classification precision was 99.88%, recall was 99.89% and F1-score 99.88%. Although it is also reasonably accurate, the performance of Logistic Regression was the lowest in all the measures of evaluation when compared to the other two models.

SVM model, the accuracy of the model was 97.86% in sentiment classification and 98.89% in the capital classification task shown in Table 3. Sentiment, precision, recall and F1- score stood 97.86%, 97.85% and 97.86%. Capital, the precision and recall were 99.93% and 99.96% respectively, F1-score 99.94%. The findings indicate a great accuracy, precision, recall, and F1-score, which makes SVM a very effective classifier in the two tasks.

2.4.2 Error Analysis via Confusion Matrices

The confusion matrices help to understand the behavior of every model better on the Sentiment and Capital classification tasks providing a complement to the aggregate evaluation measures earlier described.

The Logistic Regression has more pronounced errors in both tasks shown in figure 10 and figure 11, but it also has diverse performances with the complexity of the task. The confusion matrix used in the Sentiment task shows that there are 1,182 false positives and 918 false negatives indicating that the misclassification rate is quite high in comparison with Bi-LSTM. This is an indication that Logistic Regression is not very effective at treating non-linear association and contextual nuances, which form an important part of correct sentiment detection. Logistic Regression works far better in the Capital task where most of the predictions are along the diagonal. Nevertheless, the matrix is more dispersed than Bi-LSTM, where there is sometimes a human error between semantically related classes like Intellectual and Human Leadership Capital. These trends make it clear why Logistic Regression is a good fit to relatively linearly separable tasks, as well as why it has limited applicability to complicated and context-specific problems.

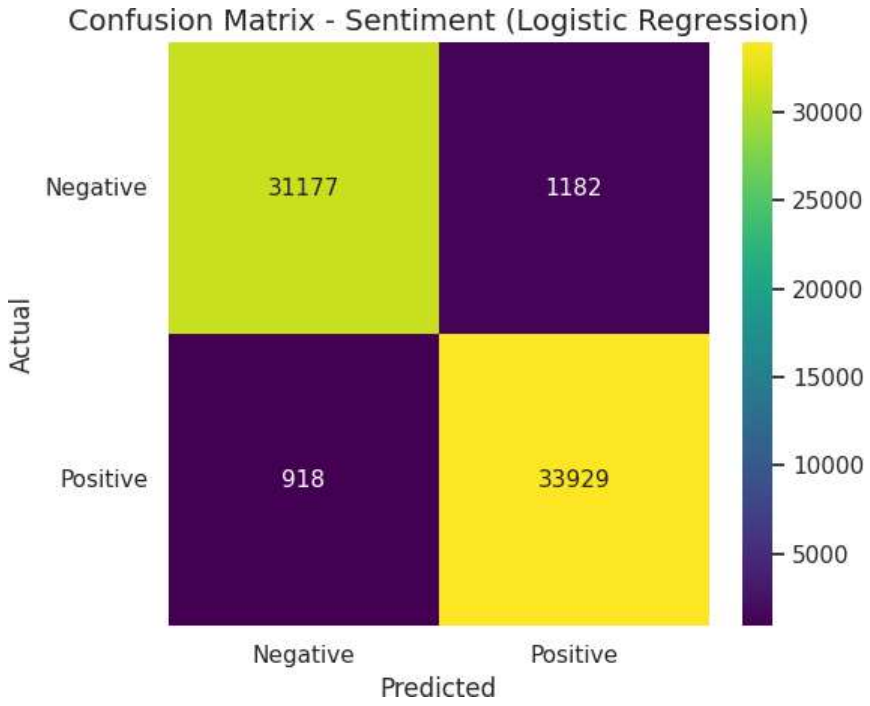

Figure 10: Logistic Regression Sentiment Confusion Matrix

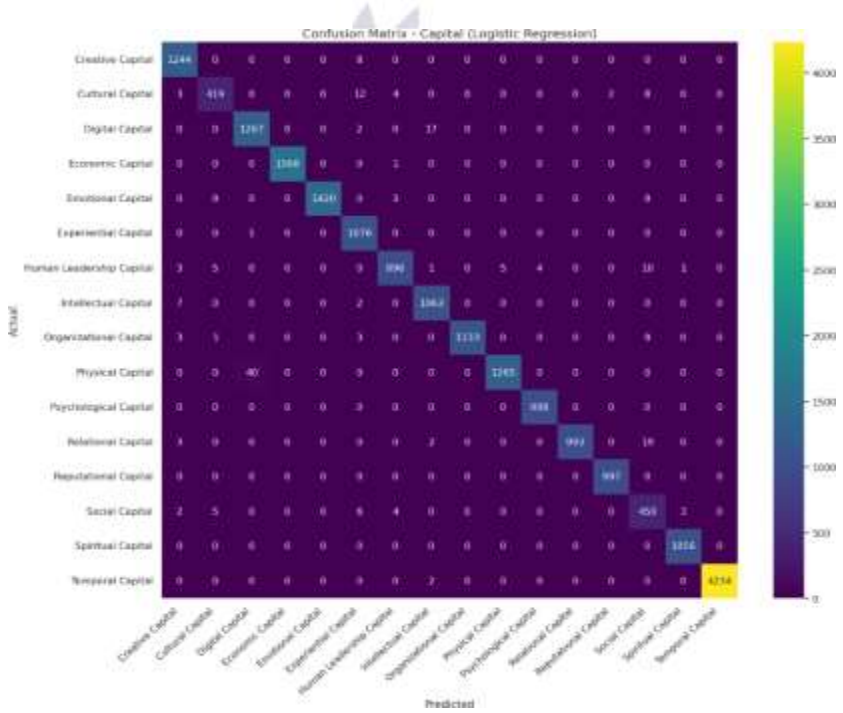

Figure 11: Logistic Regression Capital Confusion Matrix

The SVM model is a compromise between the Bi-LSTM and the Logistic Regression. In the case of the Sentiment classification in figure 12, the confusion matrix indicates that it has 756 false positives and 683 false negatives which is considerably better than the Logistic Regression but still slightly worse than Bi-LSTM. Such performance indicates the ability of SVM to learn non-linear decision borders, albeit without the

sequencing capabilities of Bi-LSTM. SVM also does well on the Capital classification task in figure 13, with the concentration of prediction along the diagonal and the mistakes few and bounded. It has similar problems as the Logistic Regression as there are sometimes off- diagonal errors in closely related categories though not as many as the Logistic Regression. All in all, SVM has shown high generalization in both binary and multi-class scenarios, and the tradeoff between accuracy and interpretability is high.

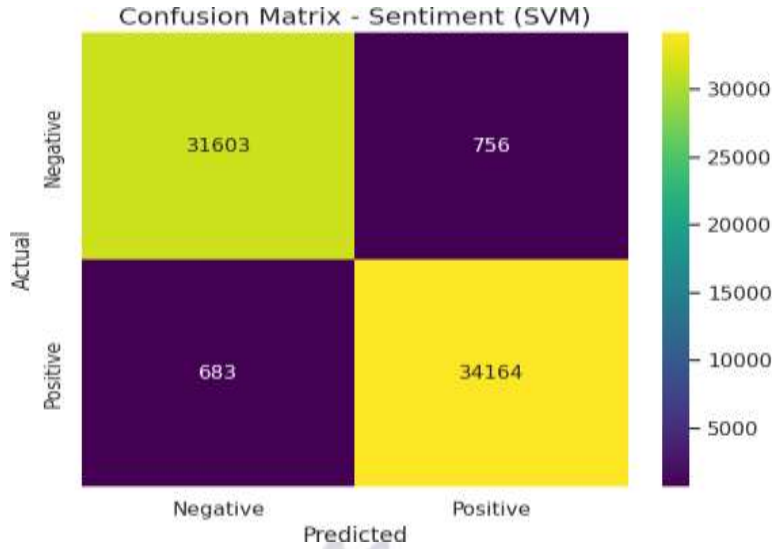

Figure 12: SVM Sentiment Confusion Matrix

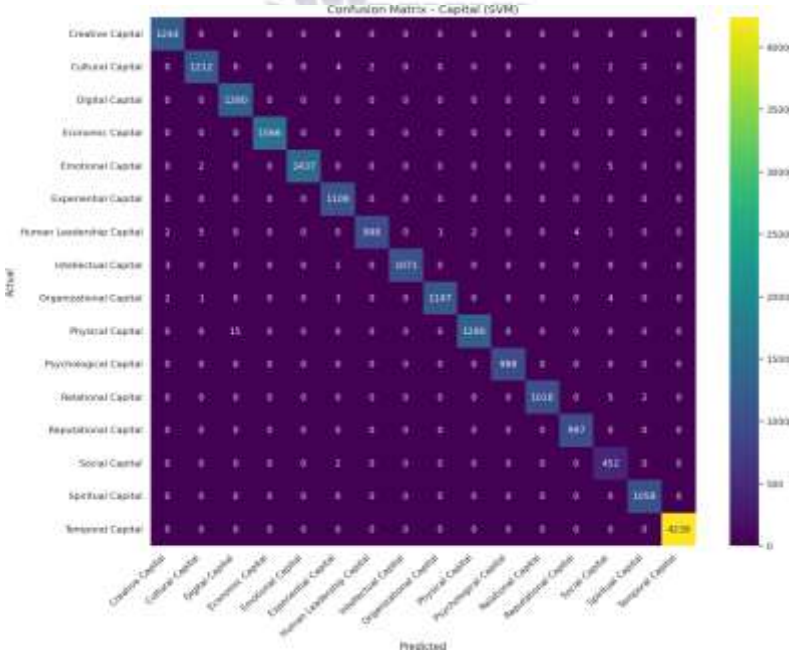

Figure 13: SVM Capital Confusion Matrix

Bi-LSTM model portrays its high-performance in both tasks. In the Sentiment classification in figure 14, the confusion matrix indicates that there is near perfect separation between the classes of data where the number of false positives and false negatives is only 172 and 223 respectively out of over 67,000 samples. This almost diagonal matrix

validates the fact that Bi-LSTM can gather contextual dependencies and subtle semantic pointers, which result in excellent precision and recall. In the Capital classification ex- periment in figure 15, Bi-LSTM also performs well with errors occurring in few sample categories. Most of the predictions can perfectly line along the diagonal, which means that the model can be able to discriminate between the fine-grained and often over-lapping categories of capitals like Cultural, Intellectual, and Human Leadership. The good suc- cess of the Bi-LSTM in binary and multi-class tasks is an indication of the benefits of the bidirectional structure of the language comprehension model in context-sensitive language understanding.

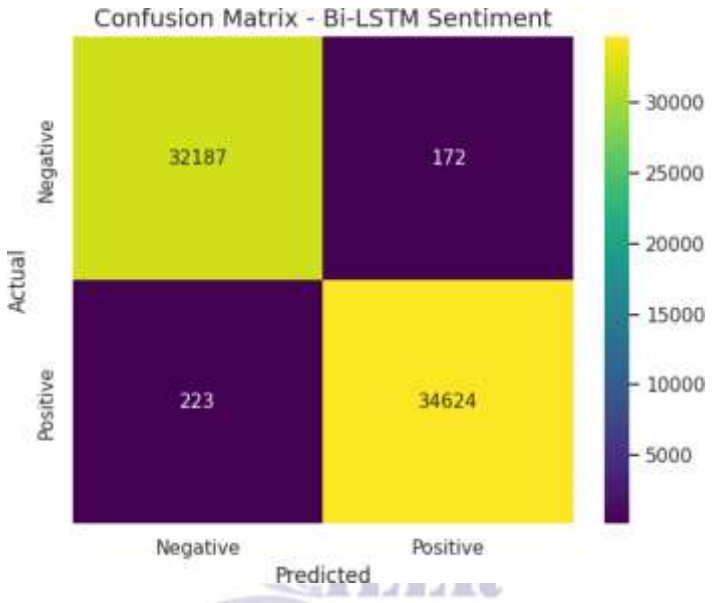

Figure 14: Bi-LSTM Multi Model Sentiment Confusion Matrix

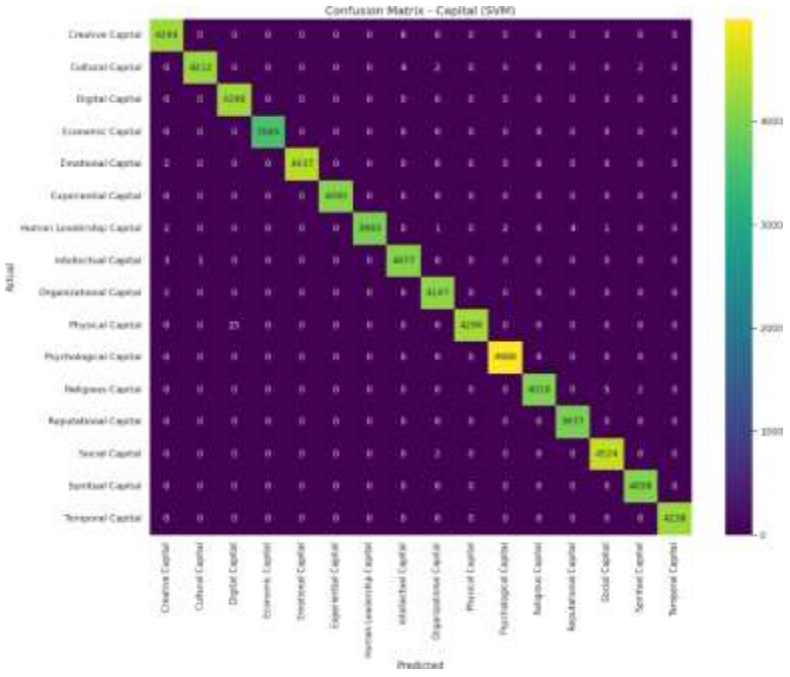

Figure 15: Bi-LSTM Capital Confusion Matrix

3 Federated Learning Results

The federated learning experiment shows that distributed training can match the level of centralized models. The global model learns efficiently, and data privacy is maintained by combining the updates of various clients. Early rounds demonstrate a high convergence rate, and the accuracy of evaluation is also constantly increasing in both sentiment and the classification of capital tasks.

3.1 Experimental Details

1. **Data Splitting:** The data is splitted into 80/20 hierarchical method that is further sub-divided to demonstrate it:

- **First Train/Test Split:** The first stage of analysis 80% of the data is used in the training process, and the others were used in the test set (20 percent).
- **Split validation:** 80 percent will be divided further into 20 percent validation and 80 percent train. It will yield a total training, validation and test split of 64, 16 and 20 percent respectively.

This guarantees the fact that the validation set and the test set are both totally unknown throughout the training and only the training set is subject to the tok- enization/Encoding to ensure the absence of data leakage. The label of the type of capital and the sentiment itself both are label-encoded relative to the training data and ultimately translated to categorical cross entropy loss.

2. **Federated learning (FL) Configurations:**

- **Local Epochs per Client:** Each client is trained on the local data for 5 epochs before sending updates to the server; these balances learning against overfitting.
- **Client Batch Size:** 32 - this has been the sweet spot to maintain stable gradient updates without overwhelming computational resource requirements.
- **Client Learning Rate:** 0.001, smaller than the centralized training to assure smooth convergence across a number of clients.
- **Number of Samples per Client:** The clients have roughly equal data par- titions, simulating realistic distributed settings.
- **Federated Rounds:** 9 global rounds of iterative aggregation of client updates into the global model.
- **Aggregation Strategy:** FedAvg used at the server to combine client model parameters while preserving data privacy.

3. **Bi-LSTM:** It is a two-output head two-output Bi-LSTM (sentiment and capital classification). Key details:

- **Vocabulary Size:** MAXWORDS= 15000 tokens, which were trained using training data.
- **Embedding:** EMBEDDIM = 128 (learned; not trained).
- **LSTM Layers:** 128-unit LSTM (Bidirectional), where the sequences are re-introduced in the pool.
- **Pooling:** Combine GlobalMaxPooling1D and GlobalAveragePooling1D and learn the strong and the average sequence level features.
- **Dropout:** 0.3 following post embedding and Bi-LSTM, 0.3-0.4 following post dense regulators.
- **Dense Layers:** 128-unit, with ReLU, and individual softmax output layers of each task.
- **Loss Function:** Custom weighted categorical crossentropy, which may permit the disequilibrium in the number of classes of the sentiment and the capital classifications.

Bi-LSTM can be used to learn sequential dependencies, whereas multi-pooling can be used to compress sequence information to provide strong classification; it is possible to design such a model.

3.2 Results

The findings of this federated learning experiment results demonstrate in Table 4, how distributed training can be compared to

centralized training and preserve privacy over the data. The pattern is similar through the first round itself: training and evaluation losses decrease very quickly, and the accuracy increases steadily. This means that although data is stored in different clients, the federated aggregation mechanism successfully attains important patterns, which facilitates the model to learn efficiently. Accuracy of distributed evaluation by Round 5 is above 93% indicating that the federated learning can reach high-quality results within relatively few rounds. The fact that the distributed and centralized measures are similar during the first rounds also indicates that there is little client drift, that is, local updates made by various clients do not create a disturbance in the overall model.

Since the training after Round 5, the distributed and centralized models show steady increases continued, the evaluation accuracy is gaining its maximum, and the loss de- creases. The distributed fit can be compared to the centralized one as it is also the case that the local updates are meaningful to global learning. The accuracy of distributed evaluation is up to 98.35 by the last round, and it is very close to the accuracy of the centralized evaluation, which is 98.59. It shows that federated learning can achieve simi- lar results as conventional centralized training without the necessity to combine sensitive client information. The pattern of the continuous enhancement also supports

Table 4: Performance Metrics Across Communication Rounds

Rounds	Distributed Train Loss	Centralized Train Loss	Distributed Fit Accuracy	Distributed Fit Loss	Distributed Eval Accuracy	Distributed Eval Loss	Centralized Eval Accuracy
0	2.7182	2.7124	0.4982	2.7182	0.5011	2.7124	0.5034
1	1.8923	1.8765	0.6123	1.6145	0.6789	1.6098	0.6821
2	1.4231	1.4109	0.7435	0.8921	0.7654	0.8876	0.7698
3	1.1056	1.0987	0.8214	0.5678	0.8342	0.5632	0.8377
4	0.8765	0.8692	0.8789	0.4123	0.8891	0.4089	0.8923
5	0.6234	0.6187	0.9123	0.2891	0.9356	0.2856	0.9389
6	0.4872	0.4821	0.9415	0.2234	0.9543	0.2198	0.9571
7	0.3989	0.3945	0.9588	0.1876	0.9672	0.1843	0.9695
8	0.3421	0.3387	0.9689	0.1623	0.9741	0.1598	0.9768
9	0.3015	0.2982	0.9756	0.1432	0.9798	0.1401	0.9823
10	0.2723	0.2691	0.9801	0.1298	0.9835	0.1267	0.9859

the stability of the federated optimization process in several rounds.

Overall, these results demonstrate the usefulness and accuracy of federated learning as a viable measure of privacy-saving AI. The little difference in performance between distributed and centralized models reveals that the federated model is adequately able to preserve the quality of the models, even under the distributed setting. The experiment helps also to highlight the necessity of attentional aggregation and learning rate policies to reduce the drift of clients and attain a smooth convergence. To sum up, this research confirms that federated learning can not only preserve the performance of a centralized counterpart but also be a powerful, scalable and secure method of training a model in a group of clients.

4 Limitations

The dataset described in this article has several limitations. It focuses solely on sentences related to specific forms of capital (e.g., Economic Capital, Social Capital), limiting its applicability to broader sentiment analysis tasks and potentially introducing biases. The dataset's narrow focus may affect the generalizability of models to other contexts. Additionally, its limited size may impact

model robustness and reliability; a more extensive, diverse dataset would likely provide more accurate results. Data collection and curation challenges may have led to an uneven sentiment distribution, influencing model performance. These limitations stem from the dataset's specialized nature and creation challenges, not from data analysis or interpretation issues.

5 Ethics Statement

The authors have adhered to the ethical requirements for publication in Data in Brief. This study uses textual data from publicly available online platforms and social-media sources. It does not contain private messages, personally identifiable or sensitive information. The data generated and curated for this research focus on sentences about various forms of capital. All ethical guidelines have been followed throughout the data collection and analysis process.

6 Conclusion

This study successfully established FedBiLSTM-Net, a novel, privacy-preserving frame- work for well-being monitoring through the integration of Federated Learning (FL) and a unified Bi-LSTM architecture. Our core achievement is the ability of our well-being meter to perform both sentiment analysis and capital prediction simultaneously. This dual-task approach is crucial because it moves beyond simply identifying an emotion to also identifying the contextual driver behind that emotion (e.g., economic or social capital), providing a far more holistic and actionable understanding of a person's state.

By utilizing FL with the FedAvg algorithm, we successfully trained the model across 20 non-IID clients, maintaining data privacy by ensuring raw text never leaves the local device. The resulting federated model achieved strong generalization, with a centralized evaluation accuracy of 80.17%. This confirms that FL is an efficient, robust and ethical mechanism for learning complex, context-sensitive patterns from diverse, distributed well- being data.

Our work delivers a collaborative, scalable, and privacy-aware solution that is highly suitable for ethically sensitive domains such as healthcare and workplace wellness. Future work will focus on expanding the model for multilingual applications, real-time monitoring and integrating personalized FL for tailored insights.

References

Banabilah, Syreen, et al. "Federated learning review: Fundamentals, enabling technologies, and future applications." *Information processing & management* 59.6 (2022): 103061.

Liu, Ji, et al. "From distributed machine learning to federated learning: A survey."

Knowledge and Information Systems 64.4 (2022): 885-917.

P. P. Shelke and A. N. Korde, "Support vector machine based word embedding and feature reduction for sentiment analysis-a study," 2020: IEEE, pp. 176-179, doi: 10.1109/ICCMC48092.2020.ICCMC-00035.

A. Zunic, P. Corcoran and I. Spasic, "Sentiment Analysis in Health and Well-being:

Systematic Review," 2020. https://medinform.jmir.org/2020/1/e16023

M. Alom, T. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. Nasrin, B. Essen, A. Awwal, and V. Asari, "The History began from AlexNet: A Comprehensive Survey on Deep Learning Approaches," 2018. https://arxiv.org/abs/1803.01164

D. Boswell, "Introduction to support vector machines," *Departement of Computer Science and Engineering University of California San Diego*, vol. 11, pp. 16-17, 2002.

T. Pranckevičius and V. Marcinkevičius, "Application of logistic regression with part- of-the-speech tagging for multi-class text classification," 2016 2016: IEEE, pp. 1-5, doi: 10.1109/AIEEE.2016.7821805.

S. T. Indra, L. Wikarsa, and R. Turang, "Using logistic regression method to classify tweets into the selected topics," 2016 2016: IEEE, pp. 385-390, doi: 10.1109/ICAC- SIS.2016.7872727.

A. Grataloup and M. Kurpicz-Briki, "A systematic survey on the application of federated learning in mental state detection and human activity recognition," 2024. https://www.frontiersin.org/journals/digital-health/articles/10.3389/fdgth.2024.1495999/full

Wasif, Dawood, et al. "Empirical Analysis of Privacy-Fairness-Accuracy Trade- offs in Federated Learning: A Step Towards Responsible AI." *arXiv preprint arXiv:2503.16233* (2025).

Mehta, Shiva, and Anubhav Bhalla. "Enhanced Sentiment Classification with Feder- ated Learning CNNs: Exploring Five Sentiment Categories." *2024 3rd International Conference for Advancement in Technology (ICONAT)*. IEEE, 2024.

Vasanthi, P., and V. Madhu Viswanatham. "Federated Learning for Multimodal Sentiment Analysis: Advancing Global Models with an Enhanced LinkNet Architecture." *IEEE Access* (2024).

Kamoji, Supriya, et al. "A Privacy Preserving Sentiment Analysis using Federated Learning." *2024 Second International Conference on Inventive Computing and Informatics (ICICI)*. IEEE, 2024.

Gholamiangonabadi, Davoud, and Katarina Grolingerauthorrefmark. "Federated Learning for Sentiment Analysis in Presence of Non-IID Data: Sensitivity of Deep Learning Models." *IEEE Access* (2024).

Pare, Tarun, Mohd Junedul Haque, and Pawan R. Bhaladhare. "Federated Learning Approach for Social Media Sentiment Analysis: Analyzing Public Opinion." *2024 8th International Conference on Computing, Communication, Control and Automation (ICCUBEA)*. IEEE, 2024.

Khan, Younas, David S ânchez, and Josep Domingo-Ferrer. "Federated learning-based natural language processing: a systematic literature review." *Artificial Intelligence Review* 57.12 (2024): 1-39.

Mohanty, Rabinarayan, Ranjan Kumar Dash, and Pradyumna Kumar Tripathy. "IFLNLP: An Improved Federated Learning and Natural Language Processing-based Smart Healthcare System." *2024 International Conference on Intelligent Computing and Sustainable Innovations in Technology (IC-SIT)*. IEEE, 2024.

Ye, Rui, et al. "Openfedllm: Training large language models on decentralized private data via federated learning." *Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining*. 2024.

Puppala, Sai, et al. "SCAN: A HealthCare Personalized ChatBot with Federated Learning Based GPT." *2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC)*. IEEE, 2024.

Peng, Yuanzhe, et al. "Multimodal Federated Learning: A Survey through the Lens of Different FL Paradigms." *arXiv preprint arXiv:2505.21792* (2025).

Huang, Wei, et al. "Fairness and accuracy in horizontal federated learning." *Information Sciences* 589 (2022): 170-185.

Riviera, Walter, Ilaria Boscolo Galazzo, and Gloria Menegaz. "FLavourite: Horizontal and Vertical Federated Learning setting comparison." *IEEE Access* (2025).

Wang, Feng, M. Cenk Gursoy, and Senem Velipasalar. "Feature-based federated transfer learning: Communication efficiency, robustness and privacy." *IEEE Trans- actions on Machine Learning in Communications and Networking* (2024).

Islam, Mynul, et al. "A federated learning approach for text classification using NLP." *PacificRim Symposium on Image and Video Technology*. Cham: Springer Interna- tional Publishing, 2022.

Hamsath Mohammed Khan, Riyas. "A Comprehensive Study on Federated Learn- ing Frameworks: Assessing Performance, Scalability, and Benchmarking with Deep Learning Model." (2023).

Shanto, Md Javed Ahmed, et al. "Federated learning empowered spam message de- tection for multilingual short message service (sms)." (2023): 1310-1311.

Sami, Abdul, et al. "Federated Learning for Sarcasm Detection: A Study of Atten- tionEnhanced BILSTM, GRU, and LSTM Architectures." *IEEE Access* (2024).

Dubey, Parul, Pushkar Dubey, and Pitshou N. Bokoro. "Federated learning for privacy- enhanced mental health prediction with multimodal data integration." *Com- puter Methods in Biomechanics and Biomedical Engineering: Imaging & Visualiza- tion* 13.1 (2025): 2509672.2021

Tran, Anh-Tu, et al. "An efficient approach for privacy preserving decentralized deep learning models based on secure multi- party computation." *Neurocomputing* 422 (2021): 245-262.

Bansal, Shefali, et al. "Federated learning approach towards sentiment analysis." *2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS)*. IEEE, 2022.

Sarracino, Francesco, et al. "A year of pandemic: Levels, changes and validity of Well-being data from Twitter. Evidence from ten countries." *PloS one* 18.2 (2023): e0275028

Eberhardt, Steffen T., et al. "Decoding emotions: Exploring the validity of sentiment analysis in psychotherapy." *Psychotherapy Research* 35.2 (2025): 174-189.

N. Zainuddin and A. Selamat, "Sentiment analysis using support vector machine," 2014: IEEE, pp. 333-337, doi: 10.1109/I4CT.2014.6914200.

L. Povoda, R. Burget, and M. K. Dutta, "Sentiment Analysis based on Sup- port Vector Machine and Big Data," 2016 2016: IEEE, pp. 543-545, doi: 10.1109/TSP.2016.7760939.

M. J. Sheller, B. Edwards, G. A. Reina, J. Martin, S. Pati, and S. Bakas, "Feder- ated learning medicine: facilitating multi- institutional collaborations without sharing patient data," 2020. https://www.nature.com/articles/s415 98-020-69250-1

Wasif, Dawood, et al. "Empirical Analysis of Privacy-Fairness-Accuracy Trade- offs in Federated Learning: A Step Towards Responsible AI." *arXiv preprint arXiv:2503.16233* (2025).

Vasanthi, P., and V. Madhu Viswanatham. "Federated Learning for Multimodal Sen- timent Analysis: Advancing Global Models with an Enhanced LinkNet Architecture." *IEEE Access* (2024).

Li, Yunbo, Jiaping Gui, and Yue Wu. "An Experimental Study of Different Aggregation Schemes in Semi- Asynchronous Federated Learning." *arXiv preprint arXiv:2405.16086* (2024).

P. Chhikara, P. Singh, R. Tekchandani, et al. "Federated Learning Meets Human Emotions: A Decentralized Framework for Human–Computer Interaction for IoT Ap- plications," 2020. https://ieeexplore.ieee.org/abstract/do cument/9253631

D. Gholamiangonabadi and K. Grolinger, "Federated learning for sentiment analysis in presence of non-IID data: Sensitivity of deep learning models", *IEEE Access*, vol. 12, pp. 12345–12360, 2024. https://ieeexplore.ieee.org/abstract/document/
10662960

C. Dwork, "Differential privacy: A survey of results", 2008. https://link.springer.com/chapter/10.1007/978-3-540-79228-4_1

K. Bonawitz et al., "Practical secure aggregation for privacy-preserving machine learning", 2017. https://dl.acm.org/doi/10.1145/3133956.3133982

H Thakur, J Srivastava, "Sarcasm Detection with BiLSTM Multi head Attention", 2024. https://ieeexplore.ieee.org/abstract/document/10543816/

W. Felbo et al., "Using millions of emoji occurrences to learn any-domain represen- tations for detecting sentiment, emotion and sarcasm," in *Proc. EMNLP*, 2017, pp. 1615–1625. https://aclanthology.org/D17-1169.pdf

A. Khan et al., "Sentiment analysis of emoji fused reviews using machine learning and BERT", 2025. https://www.nature.com/articles/s41598-025-92286-0

J Melcher, R Hays, J Torous, "Digital phenotyping for mental health of college students: a clinical review", 2020. https://mentalhealth.bmj.com/content/23/ 4/161

P. Kairouz et al., "Advances and open problems in federated learning", 2021. https: //www.nowpublishers.com/article/Details/MAL-083