INTERNSHIP SUMMARY REPORT

CDA 571TUT INT1

NIHITH NATH KANDIKATTU

UB ID : 50537232

Company: EITACIES INC

Internship Position: Machine Learning Engineer Intern

Project Name: CONFSEC

Reporting Manager: K. Akash Dharshan (kdharshan@eitacies.com)

Duration: 30th May 2024 – 16th AUGUST 2024

My Project GIT repository: https://github.com/nihith-nath/Eitacies_ML_internship_project

SUBMITTED TO Prof Rachel Hagemann Blair

December 15th Aug, 2024

University at Buffalo
The State University of New York

INTERNSHIP OVERVIEW

1. Introduction:

During my summer internship at EITACIES INC, I served as **a Machine Learning Engineer Intern**. My primary role was to develop machine learning models designed to enhance the security and integrity of video conference applications such as Zoom, Webex, Teams. The project aimed to create a robust system capable of detecting and managing risks in real-time, ensuring the protection of sensitive information and the prevention of inappropriate content during virtual meetings. Throughout the internship, I collaborated closely with my team leads and fellow interns to achieve these objectives.

2.Goals & Objective :

The primary objective of my internship was to develop machine learning models that detect and flag potential risks in virtual meetings. This includes identifying and flagging content related to PCI DSS, racial discrimination, human facial expressions, and gender discrimination. The project was designed to ensure that video conference applications could operate securely, maintaining the integrity of meetings by preventing the sharing of non-proprietary or prohibited content.

3. My Responsibilities :

My responsibilities during the internship included a range of tasks aimed at developing and deploying machine learning models. Key responsibilities included:

- **Understanding Business Objectives:** I focused on understanding the business goals related to safeguarding video conferences and developed models that aligned with these objectives. Metrics were established to track the progress of these models.
- **Algorithm Analysis:** I analyzed various ML algorithms to determine their potential success in solving the identified problems, ranking them based on their effectiveness.
- **Data Exploration and Visualization:** I explored and visualized the data to gain a comprehensive understanding, identifying differences in data distribution that could impact model performance when deployed in real-world scenarios.
- **Data Quality Assurance:** I verified the quality of the data, performing data cleaning where necessary to ensure the accuracy and reliability of the models.When additional data was required, I worked data acquisition process like creating datasets based on sample data, ensuring that the necessary data was obtained efficiently.

- **Validation Strategy Definition:** I defined validation strategies to assess the effectiveness of the models.
- **Preprocessing and Feature Engineering:** I defined the preprocessing steps and feature engineering techniques required for each dataset, as well as data augmentation pipelines to enhance model performance.
- **Model Training and Hyperparameter Tuning:** I trained the models and fine-tuned their hyperparameters to optimize performance.
- **Error Analysis and Strategy Design:** I analyzed the errors encountered by the models and designed strategies to address these issues.
- **Model Deployment:** Finally, I was responsible for deploying the models into the product environment, ensuring they operated effectively in real-world applications.

4. My Day-to-Day Activities

My daily activities were structured to ensure consistent progress on the project. These activities included:

- **Standup Calls:** Participating in daily standup calls to discuss ongoing tasks, challenges, and next steps with my team leads, **Pavan Usirika and Kusumanjali Pulivarthi.**
- **Feedback Sessions:** Regularly connecting with my team leads and my manager Akash Darshan for feedback on my work, allowing for continuous improvement and alignment with project goals.
- **Collaboration with Interns:** Engaging in brainstorming sessions with other interns, where we provided constructive criticism and shared ideas to enhance each other's work.
- **Document Review:** Reading and interpreting project documentation to ensure that my work aligned with the overall project requirements.
- **Model Development and Testing:** Implementing, testing, and refining machine learning models for various tasks, including PCI DSS detection, racial discrimination detection, facial expression analysis, and gender discrimination detection.
- **Reporting and Documentation:** Documenting my progress and the results of my work in github, ensuring that all developments were recorded and communicated to the team.

My Project GIT repository: https://github.com/nihith-nath/Eitacies_ML_internship_project

5. END-TO-END ARCHITECHTURE OF CONFSEC

Our project's architecture is designed for secure and efficient data processing. In the private zone, our machine learning analytics engine processes data from the cloud and stores results in MongoDB. This data is then accessed by our backend API in the authorized zone, which securely interacts with the user interface through role-based access control (RBAC). This layered approach ensures data integrity and security while providing users with seamless and controlled access to the insights.

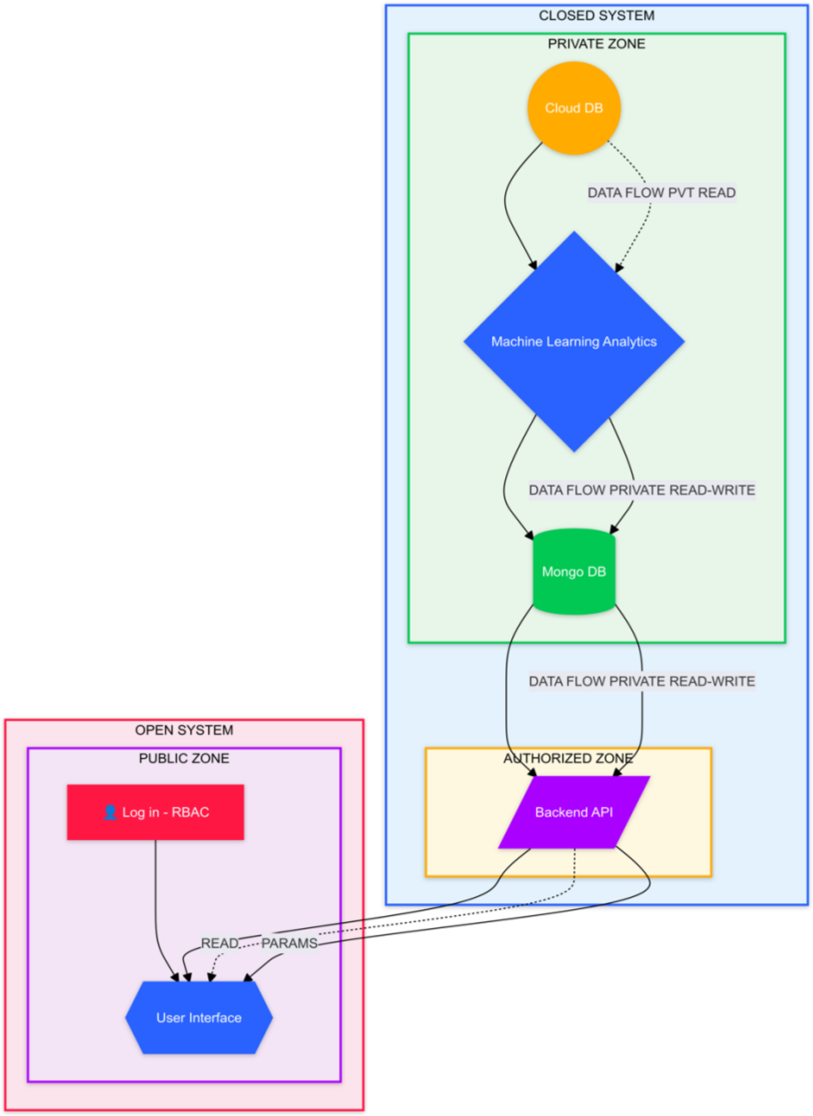


Fig: End-to-End Architecture of CONFSEC Project

TASK 1: PCI DSS Detection Using Text

Overview:
In this Task I aimed to detect sensitive information related to PCI DSS compliance in text data from virtual meetings, ensuring the prevention of unauthorized sharing of payment card information during video conference chats.

Approach:

- **Techniques Used:**
  I applied advanced text vectorization methods, including custom TF-IDF implementations, to convert text into numerical features specifically tailored to highlight potential PCI DSS violations.
- **ML Model:**
  I deployed models using Scikit-Learn, such as Logistic Regression and Decision Trees, with custom preprocessing steps that I designed to enhance detection accuracy.
- **Output:**
  The output was structured in JSON format, and I customized the input and output processes to integrate with MongoDB collections for real-time analysis and retrieval.

Statistics:

- Accuracy: 98%
- False Positives: 5
- False Negatives: 3

TASK 2: Racial Discrimination Detection

Overview:
I developed a model to identify and flag instances of racial discrimination in video conference conversations, ensuring a respectful and inclusive environment.

Approach:

- **Techniques Used:**
  I utilized advanced NLP techniques like custom word embeddings and sentiment analysis, incorporating additional preprocessing methods to capture subtle racial biases.
- **ML Model:**
  I implemented models using Support Vector Machines, integrating custom feature engineering functions that I developed to improve discrimination detection.
- **Output:**
  I formatted the results in JSON, including specific identifiers for detected racial biases, and stored them in MongoDB collections for seamless backend system integration.

Statistics:

- Accuracy: 94.44%
- Detection Rate: 97.83%
- False Negatives: 2 (Instances of racial discrimination missed)
- Bias Detection Count: 46 (True positives)

TASK 3: Facial Expression Analysis

Overview:
I focused on analyzing human facial expressions to detect emotions and assess their correlation with inappropriate behavior during meetings.

Approach:

- **Techniques Used:**
  I employed computer vision techniques and facial recognition models, enhancing them with custom preprocessing algorithms for more accurate emotion detection.
- **ML Model:**
  I used Convolutional Neural Networks (CNNs) to process and classify facial expressions, incorporating custom layers that I designed for improved feature extraction.
- **Output:**
  I converted the data to JSON with additional context for each detected emotion and stored it in MongoDB for seamless integration with the backend system.

Statistics:

- Accuracy: 60%
- Emotion Detection Count: 7

TASK 4: Gender Discrimination Detection

Overview:
I developed a system to identify and manage instances of gender discrimination in virtual meetings, promoting equality and preventing bias.

Approach:

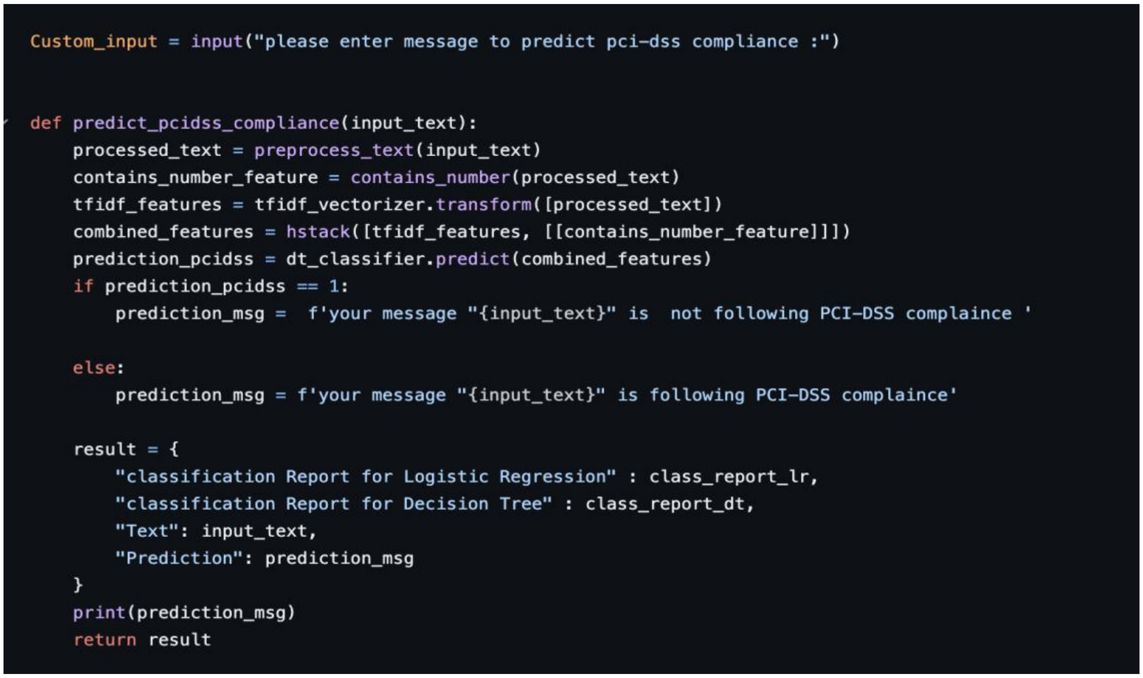
- **Techniques Used:**
  I applied NLP for gender-specific language detection and classification, leveraging BERT (Bidirectional Encoder Representations from Transformers) for deep contextual understanding and precise identification of gender biases.

- **ML Model:**
  I used BERT to capture the nuanced context of language related to gender discrimination, enhancing the accuracy and robustness of the classification process.
- **Output:**
  I stored the results in JSON format, enriching them with custom tags for detected gender biases, and imported them into MongoDB for reporting and analysis.

Statistics:

- Accuracy: 99.7%

For each task, after deploying the respective ML models, I developed custom functions that handle the entire prediction pipeline. These functions typically accept raw input (such as text or images), preprocess the data, and use the trained ML model to generate predictions. The predicted labels are then combined with the input and other relevant output data, all of which are structured in JSON format. Finally, these results are stored in MongoDB collections.



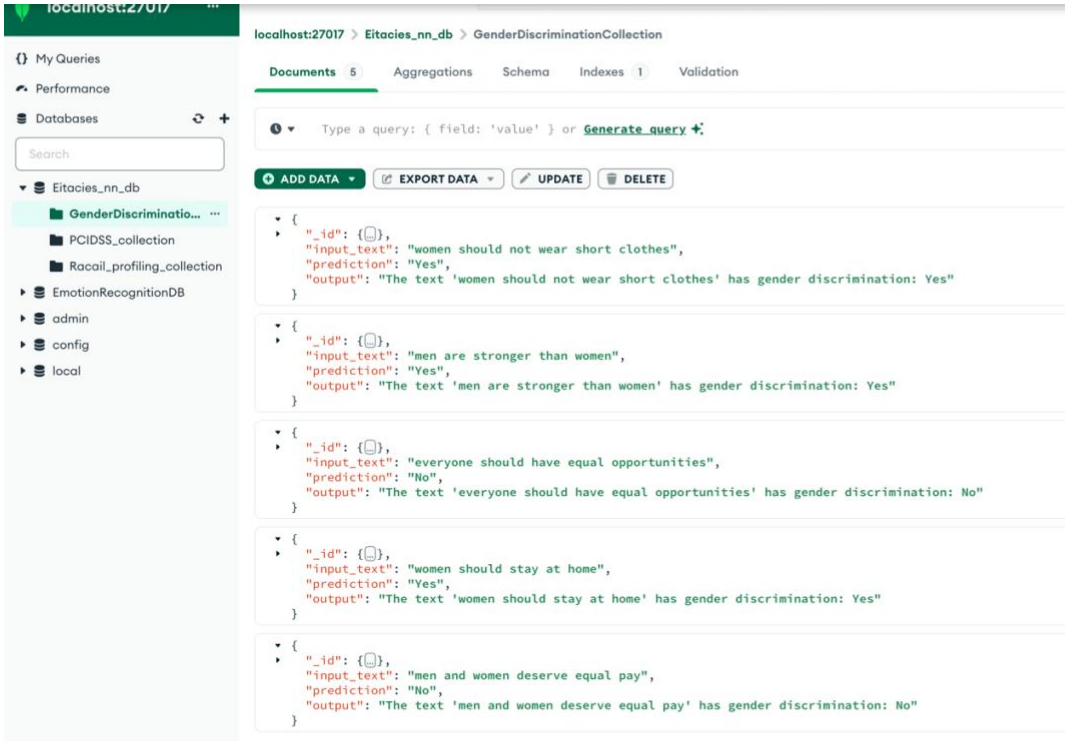
```
Custom_input = input("please enter message to predict pci-dss compliance :")


def predict_pcidss_compliance(input_text):
    processed_text = preprocess_text(input_text)
    contains_number_feature = contains_number(processed_text)
    tfidf_features = tfidf_vectorizer.transform([processed_text])
    combined_features = hstack([tfidf_features, [[contains_number_feature]]])
    prediction_pcidss = dt_classifier.predict(combined_features)
    if prediction_pcidss == 1:
        prediction_msg =  f'your message "{input_text}" is  not following PCI-DSS complaince '

    else:
        prediction_msg = f'your message "{input_text}" is following PCI-DSS complaince'

    result = {
        "classification Report for Logistic Regression" : class_report_lr,
        "classification Report for Decision Tree" : class_report_dt,
        "Text": input_text,
        "Prediction": prediction_msg
    }
    print(prediction_msg)
    return result
```

The Below figure show the deployment Results in Mongodb collections for TASK -4 Gender Discrimination other deployment images are uploaded in github



```
{
  "_id": {…},
  "input_text": "women should not wear short clothes",
  "prediction": "Yes",
  "output": "The text 'women should not wear short clothes' has gender discrimination: Yes"
}
{
  "_id": {…},
  "input_text": "men are stronger than women",
  "prediction": "Yes",
  "output": "The text 'men are stronger than women' has gender discrimination: Yes"
}
{
  "_id": {…},
  "input_text": "everyone should have equal opportunities",
  "prediction": "No",
  "output": "The text 'everyone should have equal opportunities' has gender discrimination: No"
}
{
  "_id": {…},
  "input_text": "women should stay at home",
  "prediction": "Yes",
  "output": "The text 'women should stay at home' has gender discrimination: Yes"
}
{
  "_id": {…},
  "input_text": "men and women deserve equal pay",
  "prediction": "No",
  "output": "The text 'men and women deserve equal pay' has gender discrimination: No"
}
```

6. Skills Used and Learned:

- **Machine Learning:** I applied various machine learning techniques, including Logistic Regression, Decision Trees, SVM, and CNNs, focusing on fine-tuning models for better performance.

- **NLP Techniques:** I worked with text vectorization methods like TF-IDF and developed custom word embeddings, alongside sentiment analysis and gender bias detection, to improve text data classification.

- **Data Processing:** I crafted data processing workflows that included thorough cleaning and augmentation, ensuring that the datasets were optimized for model training.

- **Computer Vision:** I implemented facial recognition and emotion detection methods, utilizing advanced computer vision techniques to analyze visual data from meetings.

- **Programming and Tools:** I used Python extensively, employing libraries such as Scikit-Learn and TensorFlow, and managed data with MongoDB and JSON. I also maintained project version control with Git/GitHub.

7.Challenges Faced and Conclusion:

Challenges Faced:

1. **Model Overfitting:** Dealt with models being overly complex, resulting in high training accuracy but poor validation performance. Overcame this by applying regularization, tuning hyperparameters, and using cross-validation.
2. **Data Collection and Validation:** Faced difficulties in obtaining high-quality, relevant data for sensitive tasks like PCI DSS and racial discrimination detection. Addressed this through synthetic data generation, data augmentation, and robust validation strategies to ensure model performance across varied datasets.

Conclusion:

My internship at EITACIES INC provided me with invaluable hands-on experience in the field of machine learning. I was able to apply theoretical knowledge to real-world problems, develop and deploy models that contribute to the security and integrity of video conferencing applications, and learn to navigate and overcome challenges in model development, data collection, and validation. The skills and insights gained during this internship have significantly enhanced my understanding of machine learning applications and have prepared me for future challenges in the field.