



Engineering
& Computing

School of Computing
& Information Sciences

School of Computing & Information Sciences

Fall 2023 Capstone II Project

Robust Intrusion Detection Systems

Students: Andres Julio-Mayorga

Mentor: **Yanzhao Wu**

Instructor: *Dr. Masoud Sadjadi*, Florida International University

Problem To Be Solved

In our interconnected digital age, the rise of malicious programs, supercharged by Artificial Intelligence (AI), poses an escalating threat to cybersecurity. The fusion of AI with cyber attacks has birthed adaptive threats capable of outsmarting traditional security systems. Intrusion Detection Systems (IDS), once stalwarts in defense, are grappling with the evolving tactics of AI-driven threats. However, the promise lies in next-generation AI-powered IDS, utilizing machine learning and deep learning to swiftly detect anomalies in real-time. As we stand at the crossroads of innovation and vulnerability, the imperative is clear: to fortify our digital defenses and foster collaboration in navigating the dynamic landscape of cybersecurity.

Our Mission

Building upon the foundation of our initial project, we undertook a progressive extension by incorporating three new cutting-edge detection models into our Intrusion Detection System (IDS). Extensive research guided the integration of these additional models, aiming to elevate the system's accuracy and fortify its capabilities against the relentless evolution of malicious programs. Our commitment to excellence led us to implement innovative preprocessing techniques, ensuring the new dataset harmonized seamlessly with the expanded ensemble of models. As we continue to evolve our project, we are steadfast in our commitment to pushing the boundaries of detection capabilities and fortifying digital defenses against the ever-evolving landscape of cyber threats.

Helpful Tools

In inheriting this project, we seamlessly continued the trajectory established by leveraging Python as our primary programming language for data processing and model training. Python, renowned for its versatility, was chosen due to its rich suite of libraries such as Numpy and Pandas for adept data manipulation, making it an inheritable asset in developing and evolving the Intrusion Detection System (IDS). The language's adaptability and ease of use enabled us to efficiently preprocess raw data, ensuring its cleanliness and preparation for ongoing model training. Furthermore, the project's foundation was strengthened by Python's extensive machine learning libraries, notably Keras and TensorFlow. This inheritance provided the essential tools to train the IDS models using a diverse set of algorithms and techniques, thereby maintaining and enhancing the system's power and accuracy.

Individual Contributions

Within the inherited project, my primary contributions centered on two fronts. Initially, I undertook a comprehensive understanding of the existing codebase, paving the way for subsequent improvements. Recognizing the importance of organizational clarity, I took the initiative to restructure the file system, enhancing overall coherence and maintainability. Simultaneously, my focus extended to research and development within the Keras module. I dedicated effort to exploring and implementing advanced models. All this, coupled with pioneering contributions to model development, played a pivotal role in advancing the project. By merging an improved file structure with better trained models, I aimed to fortify the project's foundations and ensure we approached our goals.

Summary

The evolution of the IDS project, following its adoption, stands as a testament to the prowess of integrating diverse IDS models using Python. By extending the initial framework and incorporating additional models, the final ensemble surpassed the individual models on various performance metrics. This progressive development holds significant implications for the cybersecurity industry, illuminating the potential of ensemble approaches in advancing network security and intrusion detection systems. The strategic amalgamation of multiple models not only showcased immediate success but also unveiled a promising trajectory for enhancing the overall effectiveness of these systems. In essence, this ongoing endeavor contributes valuable insights to the dynamic landscape of cybersecurity, reinforcing the project's role as a trailblazer in the field.



NumPy



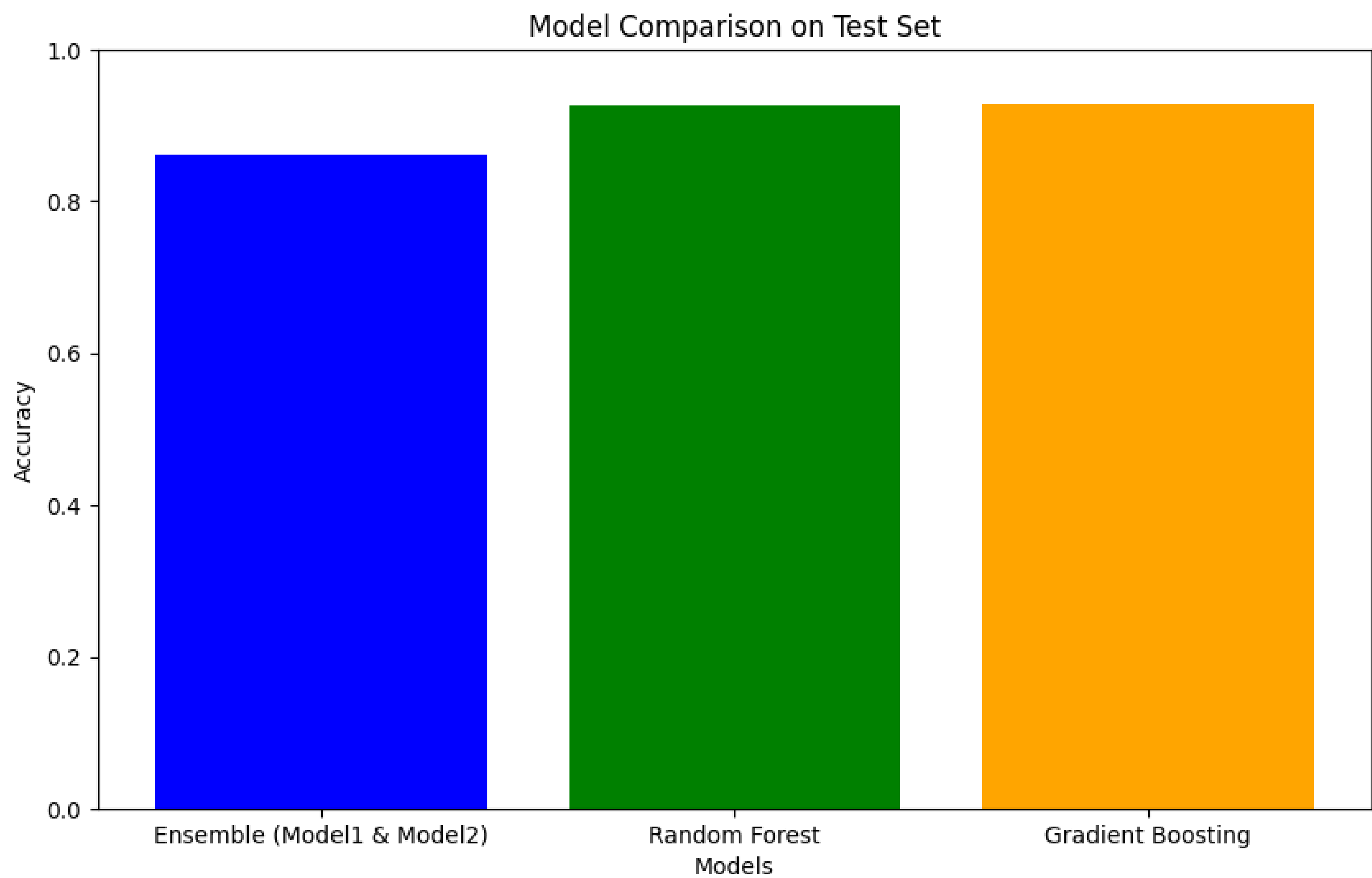
Keras

Pandas



Model Comparison

Figure 1



Navigation icons: Home, Back, Forward, Search, etc.

x=Gradient Boosting y=0.948

```
-----
Final Accuracy on Test Set (Models 1 and 2): 0.86202896
Classification Report:
      precision    recall  f1-score   support

     0       0.86       1.00       0.93       268116
     1       1.00       0.00       0.00        42913

 accuracy          0.86       0.86       0.86       311029
 macro avg          0.93       0.50       0.46       311029
 weighted avg        0.88       0.86       0.80       311029

Results for Random Forest Model:
Accuracy on Test Set (Random Forest): 0.927045388050632
Classification Report for Random Forest Model:
      precision    recall  f1-score   support

     0       0.73       1.00       0.84        60593
     1       1.00       0.91       0.95       250436

 accuracy          0.93       0.93       0.93       311029
 macro avg          0.86       0.95       0.90       311029
 weighted avg        0.95       0.93       0.93       311029

Results for Gradient Boosting Model:
Accuracy on Test Set (Gradient Boosting): 0.9289069507988001
Classification Report for Gradient Boosting Model:
      precision    recall  f1-score   support

     0       0.73       1.00       0.85        60593
     1       1.00       0.91       0.95       250436

 accuracy          0.93       0.93       0.93       311029
 macro avg          0.87       0.95       0.90       311029
 weighted avg        0.95       0.93       0.93       311029

Accuracy Difference between Models 1/2 and Random Forest Model: 0.06501643163736293
Accuracy Difference between Models 1/2 and Logistic Regression Model: 0.0668779943855311
```

Acknowledgement

The material presented in this poster is based upon the work supported by Yanzhao Wu. I thank Yanzhao Wu for assistance, cooperation and mentorship that we received throughout this process.