DDoS Attack Detection Sprint 3

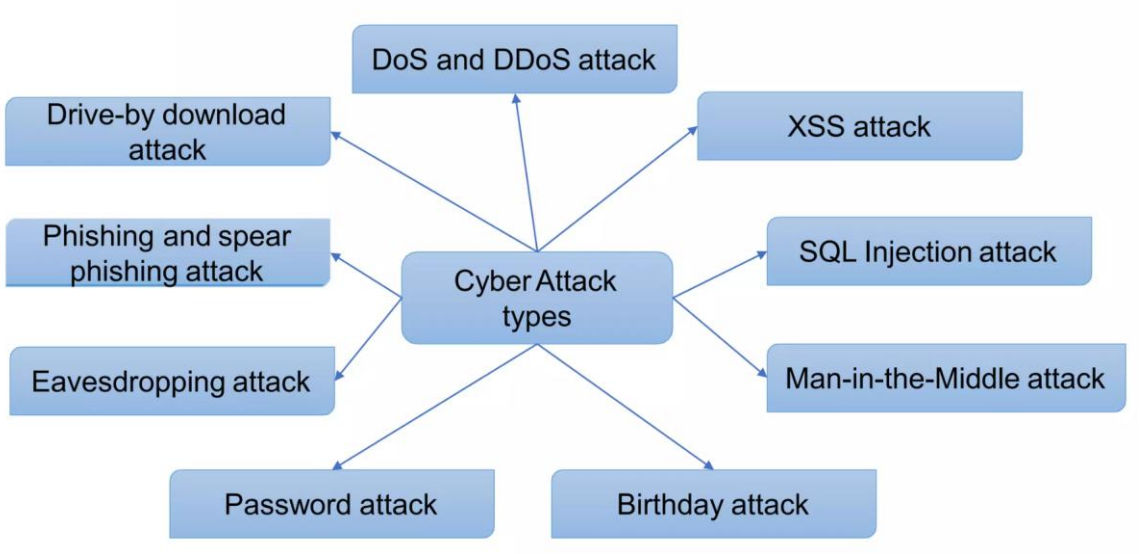
AiDoS Askhatuly

Limitations of Traditional Security Solutions

Traditional security solutions such as firewalls, antivirus software, and cryptography systems have long been the cornerstone of cybersecurity strategies.

However, these solutions are no longer sufficient to address the evolving threats in today's cyber landscape. The key challenges:

- Static Defence (e.g., relying on signatures)
- Limited Visibility (due to complexity of infrastructure)
- Lack of Advanced Threat Detection (e.g., zero-day attack)



Leveraging Data Science to Detect DDoS Attacks

DDoS attacks pose a significant threat to organizations, disrupting services and causing downtime. Data science techniques offer effective means to detect and mitigate these attacks.

Implementing data science solutions for DDoS attack detection can have a significant impact on enhancing cybersecurity defenses and mitigating the effects of DDoS attacks:

1.Improved Threat Detection. Detect DDoS attacks in real-time, minimizing the impact of these attacks on their systems and services.

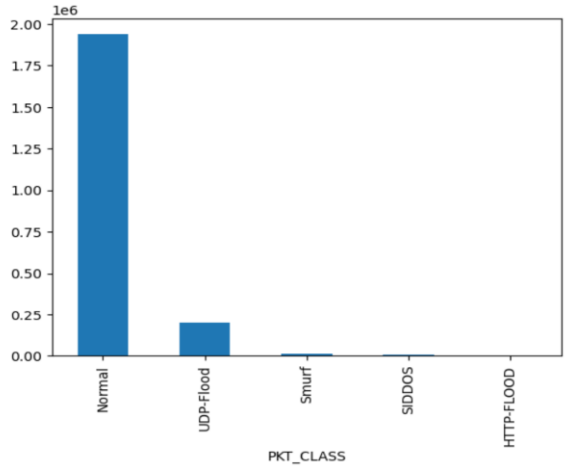
2.Reduced Downtime. Early detection of DDoS attacks enables organizations to initiate timely response measures, such as traffic rerouting or mitigation strategies, to minimize service disruptions and reduce downtime for users.

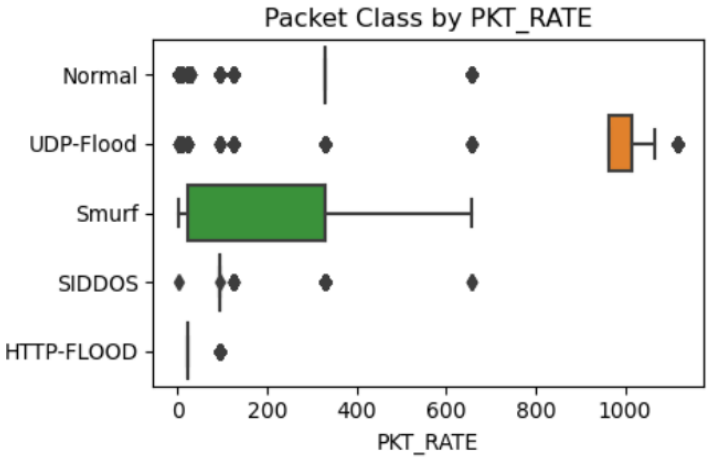
3.Enhanced Resilience. More resilient networks and systems capable of withstanding DDoS attacks, thereby maintaining continuous service availability and reliability for users.

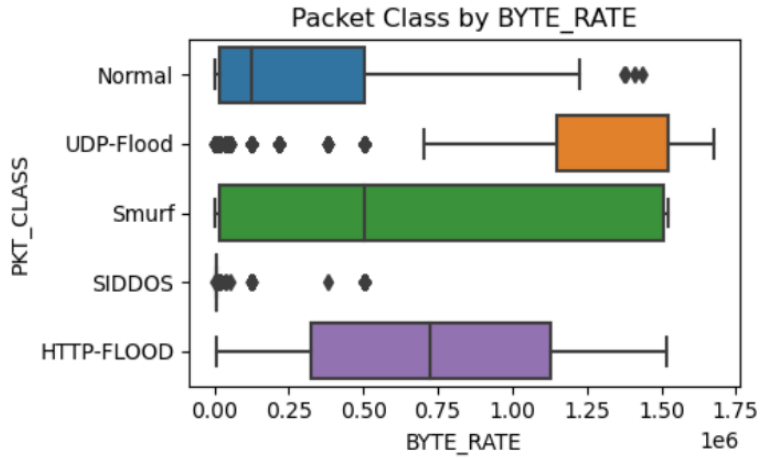
4.Cost Savings. Avoid potential revenue losses associated with downtime and service disruptions, leading to cost savings in terms of operational and reputational damage.

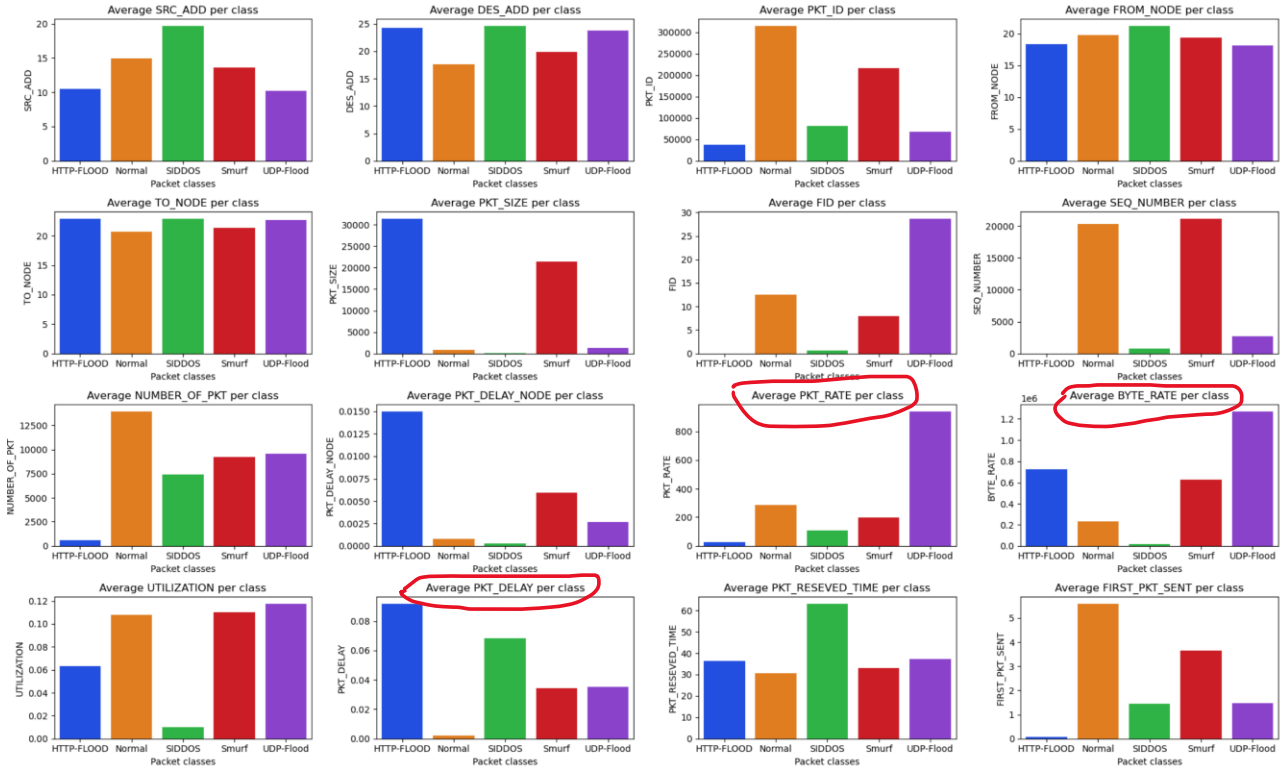
Intro to the Dataset and Preliminary EDA

- **Source:** Network traffic analyzing tool Wireshark.
- **Shape:** 2 160 668 * 28.
- **Concerns:** imbalanced data 90*10, IP address and hostnames are anonymized, collinearity, correlation between independent variables.
- **Data is clean:** no null values, no duplicates



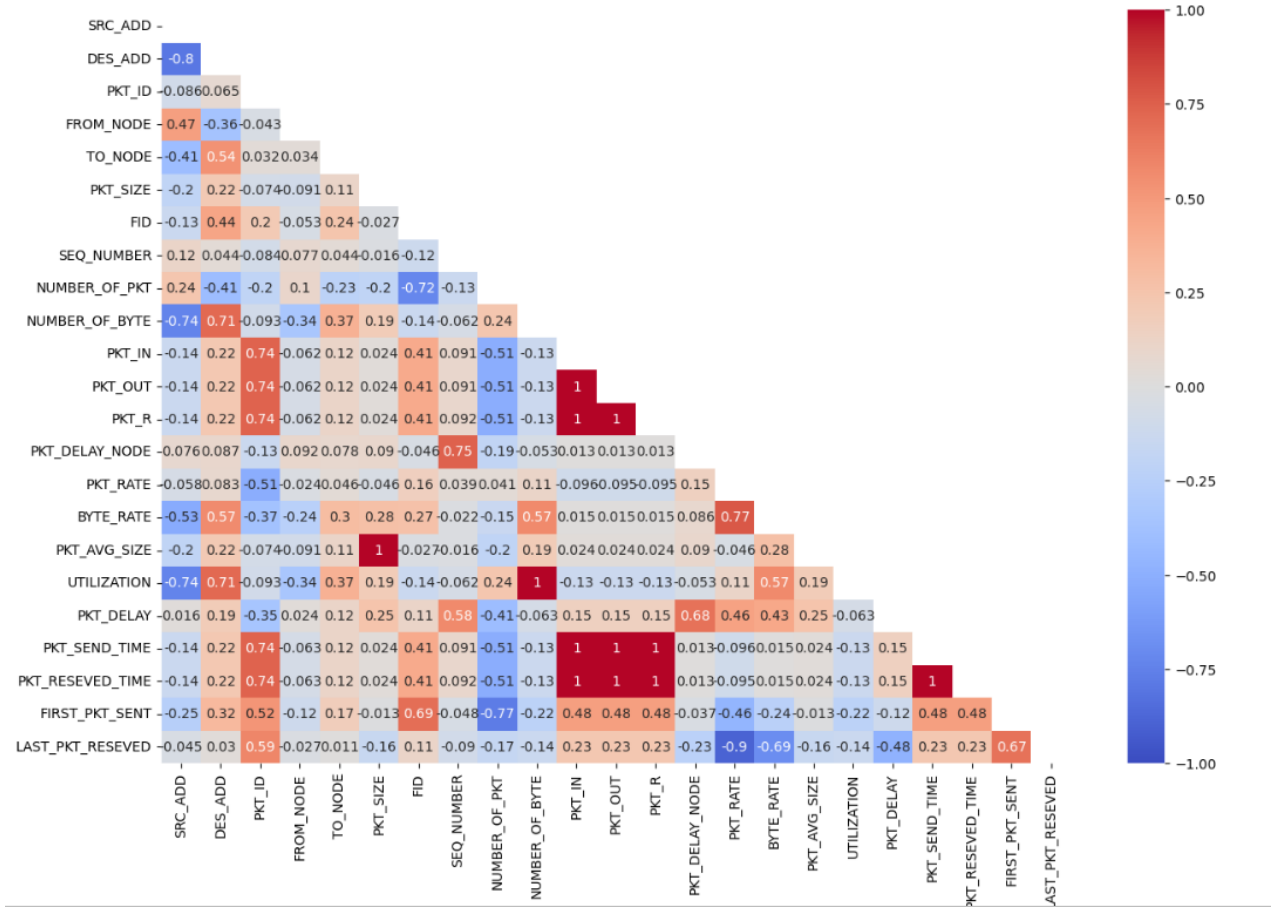




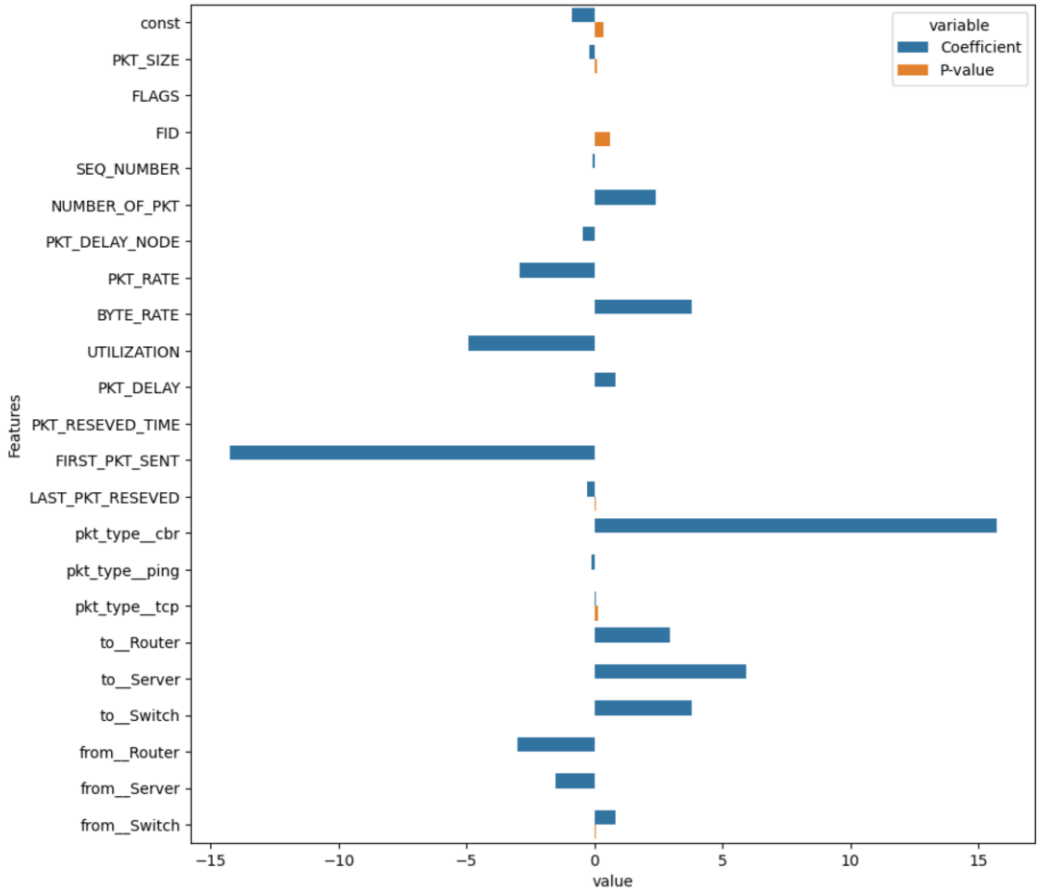


Preprocessing

- Transformed from multiclass to binary problem
- Removed collinearity
- One-hot-encoding and label encoding
- Class balancing (Under sampling, Smote)
- Scaling

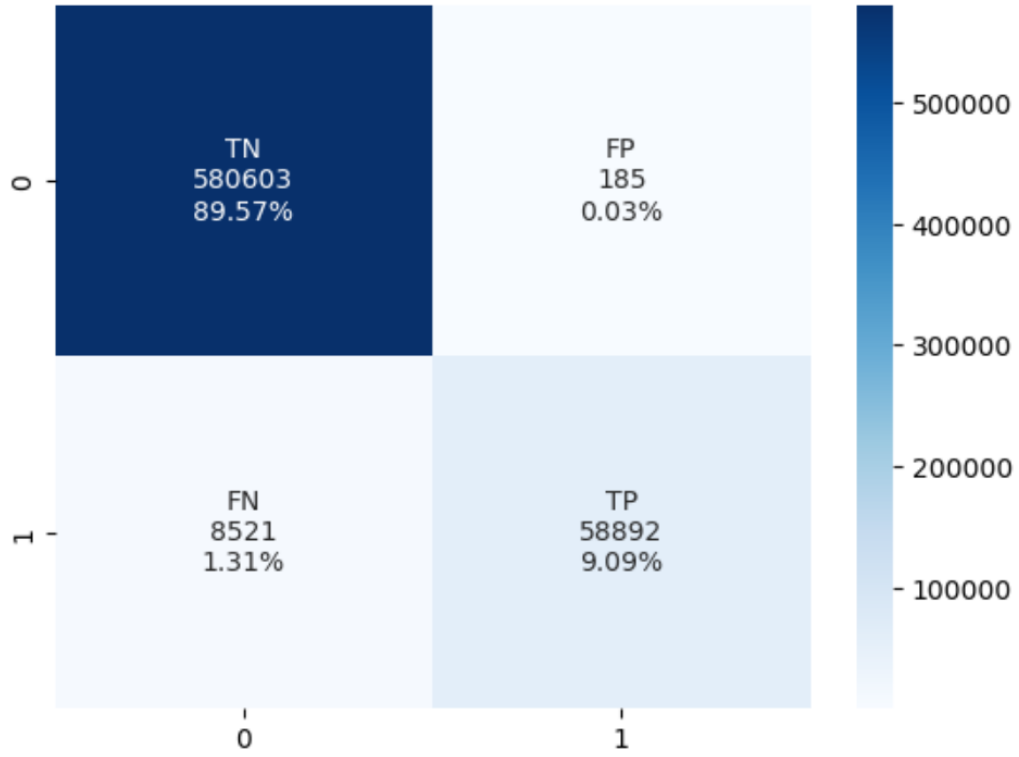


Baseline Models and Evaluation Metrics

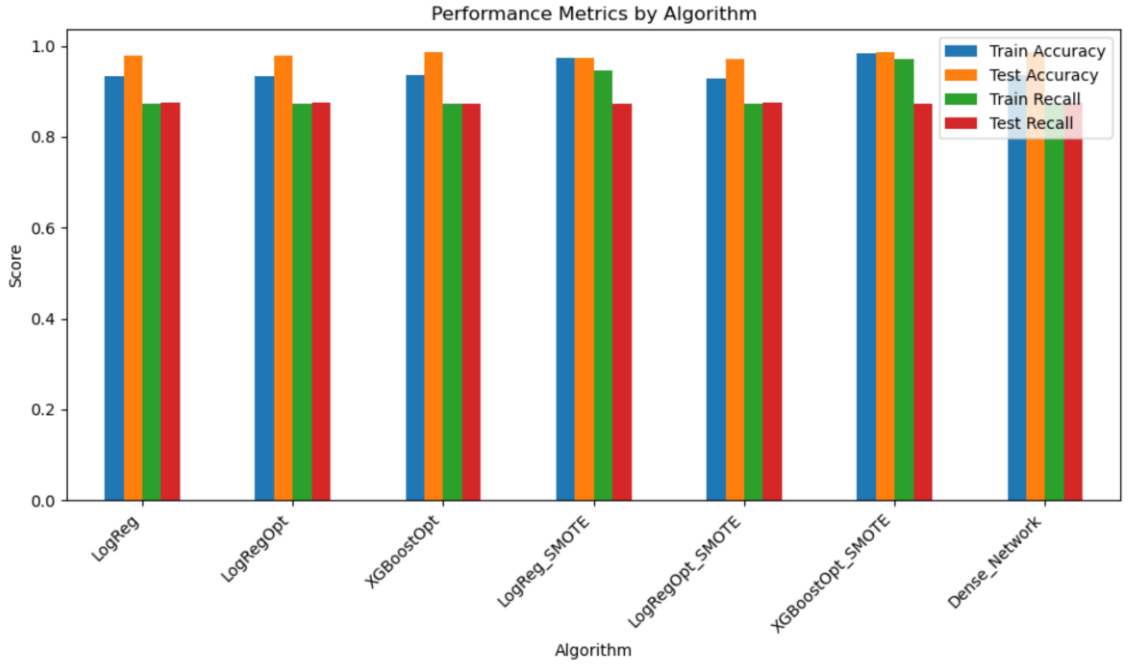


Logistic Regression performance:

- Accuracy – 98.6% (test) and 93.6% (train)
- Precision score – 99.7%
- Recall score – 87.4%
- F1 score – 93%
- AUC – 94%



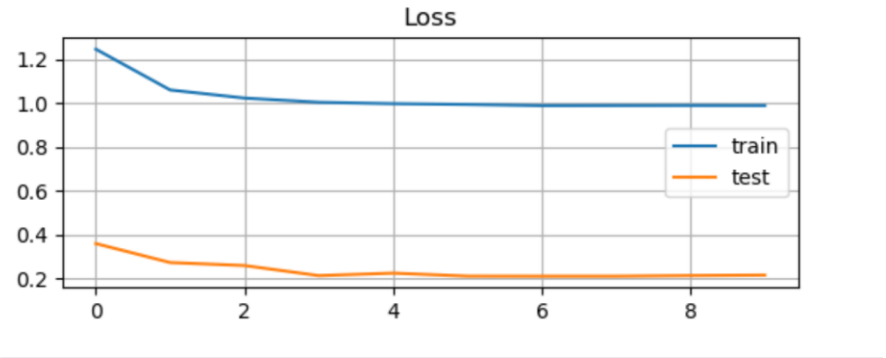
ML Models Performance

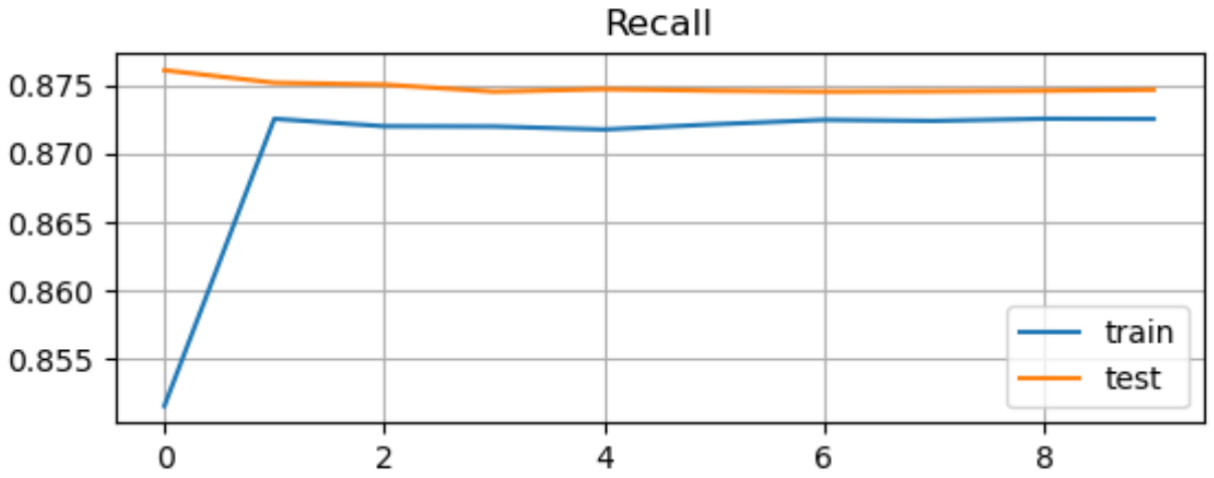


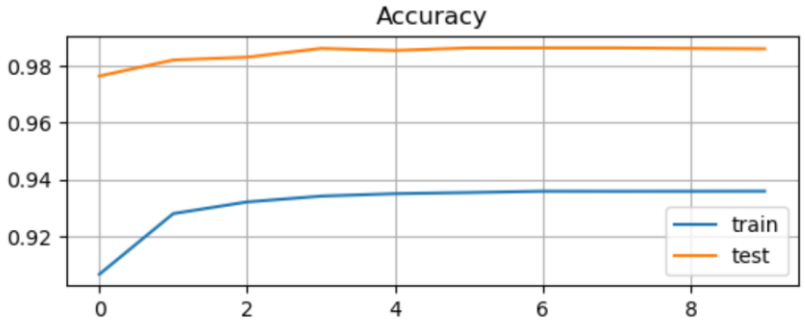
	Train Accuracy	Test Accuracy	Train Recall	Test Recall
LogReg	0.932865	0.980000	0.873570	0.874416
LogRegOpt	0.932382	0.979144	0.873640	0.874490
XGBoostOpt	0.936238	0.986731	0.872635	0.873541
LogReg_SMOTE	0.973243	0.973243	0.946838	0.873570
LogRegOpt_SMOTE	0.927627	0.970716	0.873502	0.874490
XGBoostOpt_SMOTE	0.985097	0.986712	0.970357	0.873541
Dense_Network	0.935867	0.985993	0.872513	0.874638

Neural Network Performance

- **Dense Network (5 layers*32 nodes)**







A Product Demo