## Business Requirement Document

Project Summary

Any unauthorized activity on a computer network which could threaten users' privacy or harm the function and infrastructure of the whole network is known as network intrusion. Due to the exponential growth of computer networks and web applications, it has become a critical security requirement to have the ability to protect against the potential threats that can be caused by network attacks.

The main purpose of this system is to ensure that the user is notified when an attack or intrusion takes place so as to have a quick response to malicious traffic.

Project Use Case

Use Case Diagram
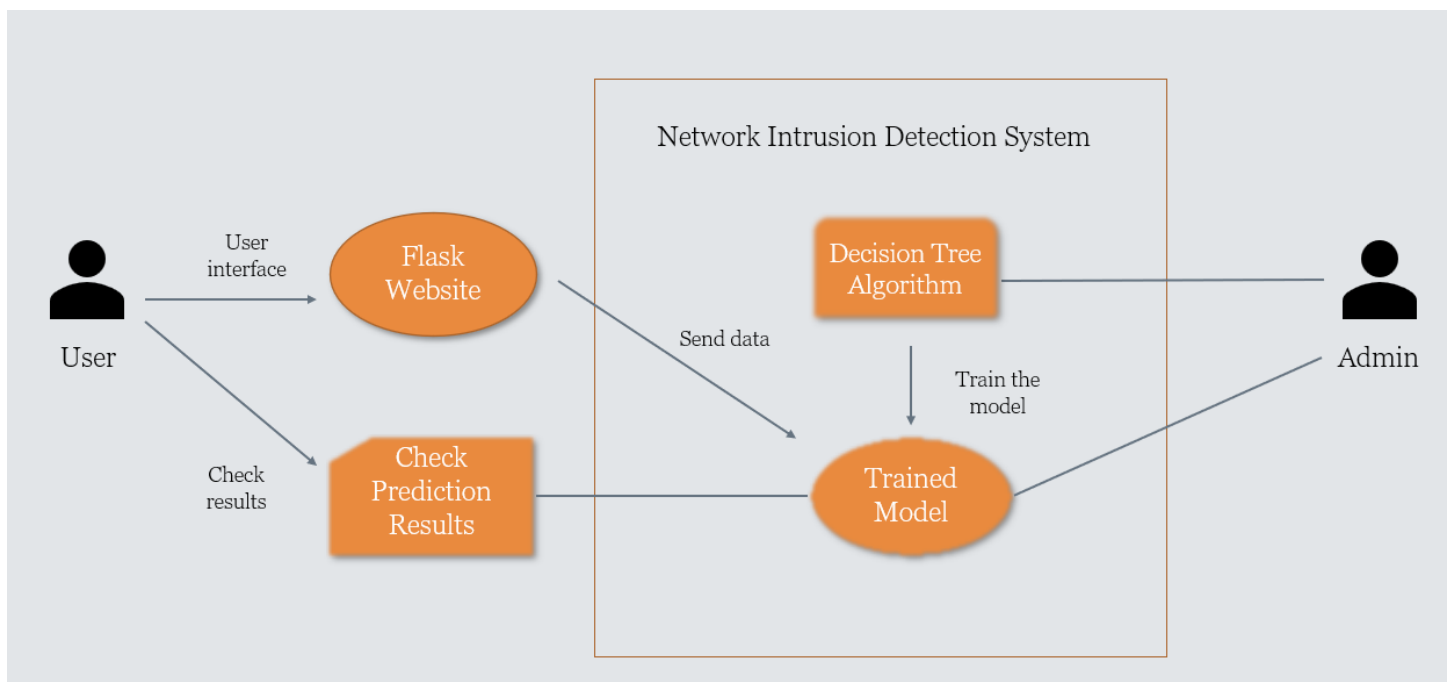
Our network intrusion system is a predictive model or a classifier which is capable of distinguishing between malicious networks also called intrusions or attacks and good normal connections. The user can check whether a network connection poses threat to their machine by just adding network parameters such as number of connections to the same destination host as the current connection in the past two seconds, status of connection, flag value, if the

user is logged in etc. and then the system will predict if the network is normal or if it is classified as an attack. It could also be used in these following applications

- Monitoring the operations of routers, firewalls, and servers.
- Recognizing and reporting when the IDS detects that data files have been altered
- Providing a user-friendly interface so non-expert staff members can assist with managing system security

Business Requirements

Priority Level	Completed	Requirement Description
Medium	Yes	Finding dataset fulfilling big data requirements.
High	Yes	Using Apache Spark to perform data extraction, cleaning, exploration, and preprocessing
High	Yes	Using Apache Spark Machine Learning Library to predict label class
High	Yes	Evaluating performance for best model
Medium	Yes	Development of website using Flask
Low	Yes	Deployment of website on Heroku

Key Stakeholders

Name	Role	Duties
Robin Huang	Professor	Grading
Shubham Chawla 200403036	Developer	Development of machine learning model using Apache Mllib, flask website and deployment to Heroku.
Archit Sinha 200505416	Business Analyst	Data Extraction, Data Cleaning, Data Exploration
Chintan Piyush Vajani 200508118	Project Manager	Drafting Business Requirement Document, Technical Design Document and Power Point Presentation.

Project Constraints

Constraint	Description
Timeline	Completing Project with all other course exams and projects
Memory Consumption	As the data set had more than 4 million rows, preprocessing and model training took more time
Team availability	Must stick to team schedules