Network Slicing Recognition

Group 3
Priya Varahan
Priyanka Akula
Shreya Chikatmarla
Sowjanya Pamulapati
Nandini Sreekumaran Nair

Agenda

- Introduction
- Motivation
- Objective
- Project Workflow
- Technology and Literature Survey
- Project Resource Requirements
- Data Collection
- Data Exploration
- Data Cleaning
- Exploratory Data Analysis
- Data Preprocessing
- Data Preparation
- Machine Learning Model
- Model Evaluation
- Model Comparison
- Ensembling
- Model Deployment
- Conclusion
- Future Scope
- References

Introduction

- Network Slicing Recognition is pivotal for the digital transformation in the telecom industry.
- 5G introduces the ability for end-to-end resource allocation through Network Slicing (NS).
- NS offers the flexibility to tailor network slices based on various parameters like bandwidth, coverage, security, latency, and reliability.
- This customization is crucial for the diverse requirements of 5G applications, use cases, and services.

Motivation

- Complexity of modern networks is resolved and enhanced user experience by using Network Slicing
- An enhanced predictive model for the telecom industry using Machine Learning to detect and mitigate network vulnerabilities is the goal.
- This model will analyze incoming connections and parameters to automatically select the best network slice, even during network failures.

Objective

- Develop a proactive ML model for 5G threat detection to help identify and eliminate threats quickly.
- Enable dynamic network slice selection to maintain service in network failures. End-to-end slice isolation protects against threats.
- Optimize performance by customizing slice configurations for service needs.
- Integrate ML with network slicing for proactive threat management and service optimization.

Project Workflow

Business Understanding

-Establish objectives through team meetings and brainstorming.
-Analyze the Problem Statement.
-Develop a Data Mining Plan.

Data Understanding

-Gather Dataset
-Execute EDA
-Perform Data Quality checks

Data Preparation

-Data pre-Processing
-Feature selection for SVM, Random Forest, Logistic Regress, KNN, Decision Tree

Data Modelling

-Datasets are divided into two groups train and test sets.
-Using the train dataset, we are running the model pipeline, which has base models, auto-ML modeling and ensemble techniques.

Data Evaluation

-Evaluate the model and visualize the results.

Deployment

-Deploying results onto dashboard and into Github repository

Technology and Literature Survey

Title	Dataset	Models	Results
5G Network Slicing: Analysis of Multiple Machine Learning Classifiers	Iranian agriculture and livestock production data from 1961 to 2017 from FAOSTAT.	Adaptive network-based fuzzy inference system (ANFIS),MLP	A thorough examination of logistic regression, linear discriminant, kNN, decision tree, random forest, SVC, BernoulliNB, and GaussianNB models for 5G network slice detection..
Harris Hawks optimization based hybrid deep learning model for efficient network slicing in 5G network	Unicauca IP flow version2, 5G network slicing dataset	CNN and LSTM models	We show that the HHO-CNN+LSTM outperforms many performance parameters on the Unicauca and 5G Network Slicing datasets.
Machine Learning-Based Network Sub-Slicing Framework in a Sustainable 5G Environment	describes linked devices, their application requirements, network resources, and performance data.	SVM K-means	Virtualized sub-slice division improves network load balance and power efficiency, improving performance and energy efficiency. Framework optimization and security assessment are planned.

Technology and Literature Survey

Title	Dataset	Models	Results
Integration of Network Slicing and Machine Learning into Edge Networks for Low-Latency Services in 5G and beyond Systems	Mobile Core data	RL Algorithm	Addressing edge-enabled network slicing synchronization issues, delivering a new system architecture, data utilization strategy, and ML integration to improve 5G and beyond QoS, QoE, and scalability. Syncing hierarchically distributed SDN controllers with RL.
Traffic analysis for 5G network slice based on machine learning	The 11 traffic types involved are WWW, FTP, DATABASE, P2P, SERVICE, MAIL ATTACK	LassoCV, classification	The research proposes a family traffic analysis system that uses the Internet of Things to accurately identify malware threats. The implemented model's accuracy is nearly 100%, making it a good traffic classification reference.
ADAPTIVE6G: Adaptive Resource Management for Network Slicing Architectures in Current 5G and Future 6G Systems	dataset containing SMS, call, and historical Internet records	traditional ML Algorithms	The ADAPTIVE6G framework for B5G and 6G systems optimises network slicing for better resource management, performance, and error reduction, while Transfer Learning (TL) using a pre-trained model yields faster, more accurate Energy efficiency, privacy, and traffic prediction and optimization analytics are future targets.

Project Resource Requirements

Hardware Requirements			
Resources	**Configuration**	**Purpose**	**Cost**
Local Machine	Chip featuring an 7-core CPU, 7-core GPU, 4 performance cores and 4 efficiency cores 2TB SSD, 16-core Neural Engine, with 64GB RAM	Processor required for Jupyter notebooks,Visual Studio to run Deep learning models	1600$

Project Resource Requirements

Tools and Licenses			
	Purpose	**Licenses**	**Cost**
Jupyter Notebook	Code development	Proprietary	Free
GitHub	Via web version system for oversight	Multiple	Free
Google Docs, slides	Word processing, spreadsheet, and presentation applications	Proprietary	Free
Discord, WhatsApp, Email	For team meetings	Free	Free

Project Libraries and Packages Requirements

Libraries and Packages Requirements			
	Library	Method	Usage
Pandas	pandas pandas,plotting	DataFrame, Series autocorrelation plot	visualization, manipulation, and time series analysis
Matplotlib	Matplotlib.pyplot	pyplot	Used for data visualization and plotting
numpy	numpy	numpy.array()	Image pixels as matrices and mathematical operations
Scikit-Learn	Sklearn.metrics	Mean_squared_error	Used for model evaluation process
Seaborn	Seaborn	sns.heatmap() sns.pairplot()	Enhances Matplotlib plots, simplifies visualization creation

Data Collection

- Our dataset was obtained initially from the 'Data Sprint 86 Network Slicing Recognition' challenge on the AI Planet platform.
- It was originally put together for the Network Slicing Recognition dataset Challenge which is a chance for students to conduct research or analysis on network data and share their discoveries.

Dataset: https://aiplanet.com/challenges/254/data-sprint-86-network-slicing-recognition-254/data

Data Exploration

Dataset Overview

LTE/5g Category	Time	Packet Loss Rate	Packet delay	IoT	LTE/5G	GBR	Non-GBR	AR/VR/Gaming	Healthcare	Industry 4.0	IoT Devices	Public Safety	Smart City & Home	Smart Transportation	Smartphone	slice Type
14	0	0.000001	10	1	0	0	1	0	0	0	0	1	0	0	0	3
18	20	0.001000	100	0	1	1	0	1	0	0	0	0	0	0	0	1
17	14	0.000001	300	0	1	0	1	0	0	0	0	0	0	0	1	1
3	17	0.010000	100	0	1	0	1	0	0	0	0	0	0	0	1	1
9	4	0.010000	50	1	0	0	1	0	0	0	0	0	1	0	0	2

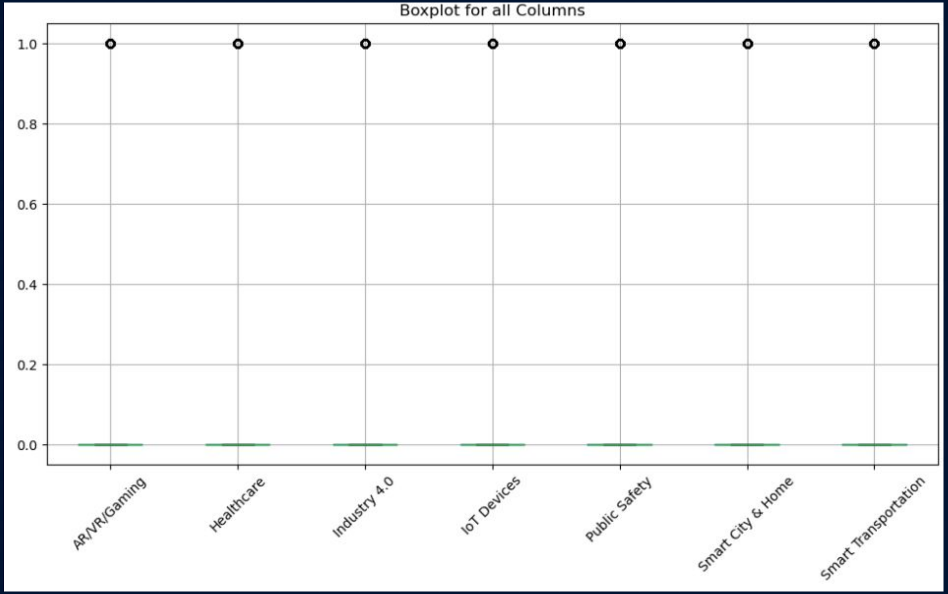
- LTE/5G - User Equipment categories or classes defining performance specifications for network devices.
- Packet Loss Rate - The ratio of lost packets to the total packets sent, indicating network stability and reliability.
- Packet Delay - The time taken for a packet to be received, an essential metric for assessing network responsiveness.
- Slice Type - Configurations allowing multiple virtualized and independent networks, facilitating service customization.
- GBR (Guaranteed Bit Rate) - Ensures a minimum rate of data transfer, crucial for maintaining service quality.
- Healthcare, Industry 4.0, IoT Devices, Public Safety, Smart City & Home, Smart Transportation - Binary indicators (1 or 0) of usage in respective domains.
- Smartphone - Indicates whether the network slice is used for cellular data on smartphones.

Data Cleaning

Checking the NULL values

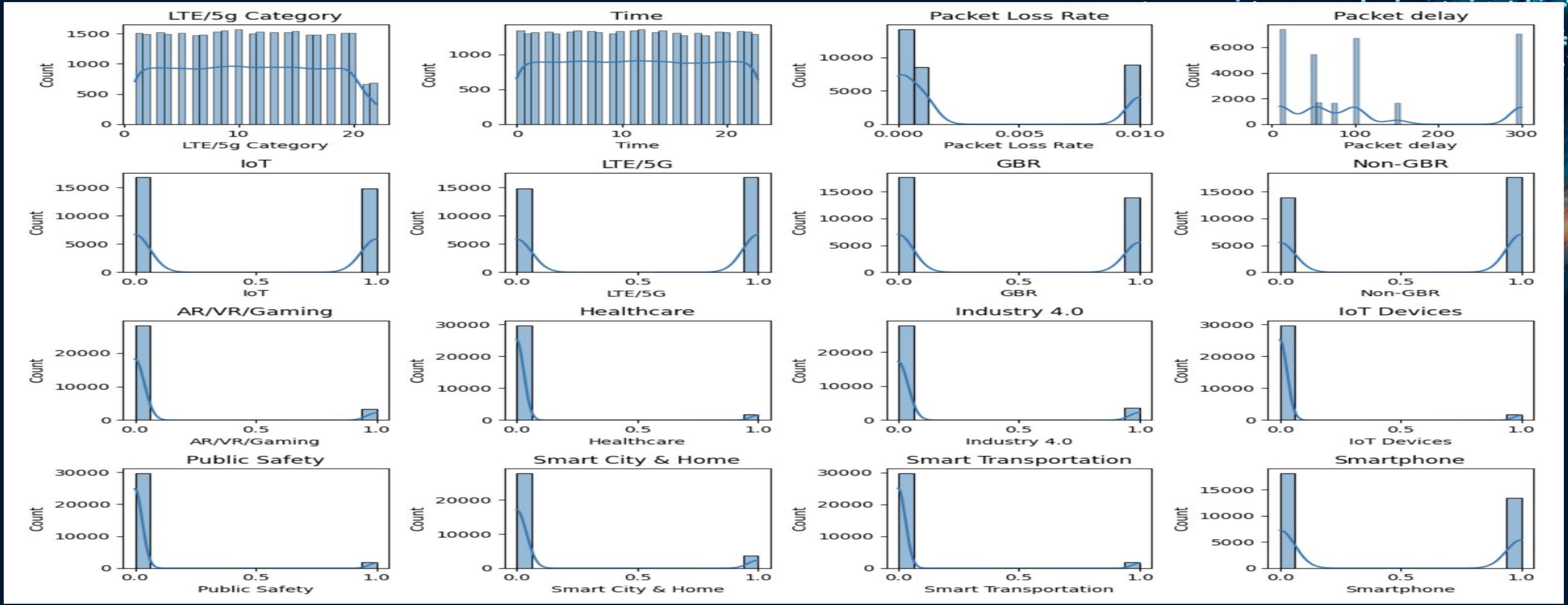
```
df.isnull().sum()

LTE/5g Category            0
Time                       0
Packet Loss Rate           0
Packet delay               0
IoT                        0
LTE/5G                     0
GBR                        0
Non-GBR                    0
AR/VR/Gaming               0
Healthcare                 0
Industry 4.0               0
IoT Devices                0
Public Safety              0
Smart City & Home          0
Smart Transportation       0
Smartphone                 0
slice Type                 0
dtype: int64
```

Boxplot of columns with outliers (based on IQR score)



Exploratory Data Analysis

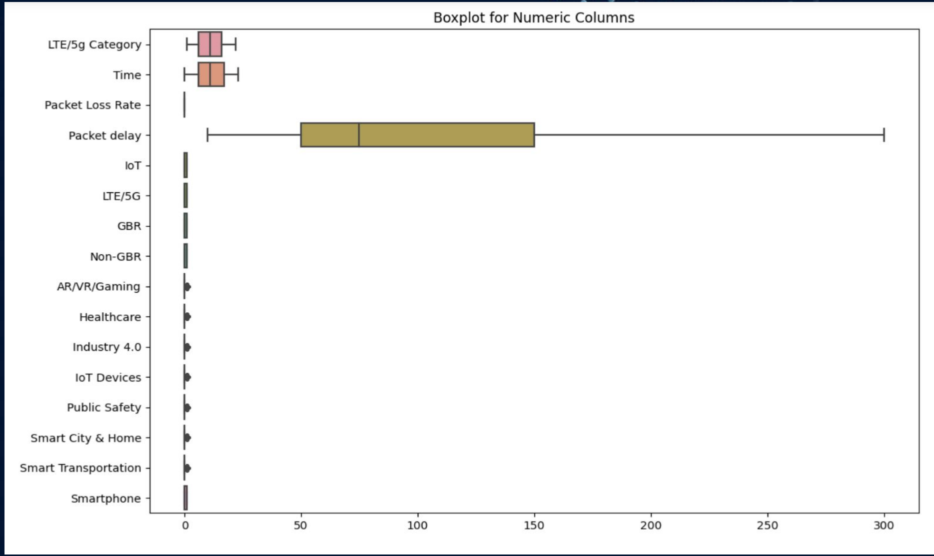
Distribution of feature values



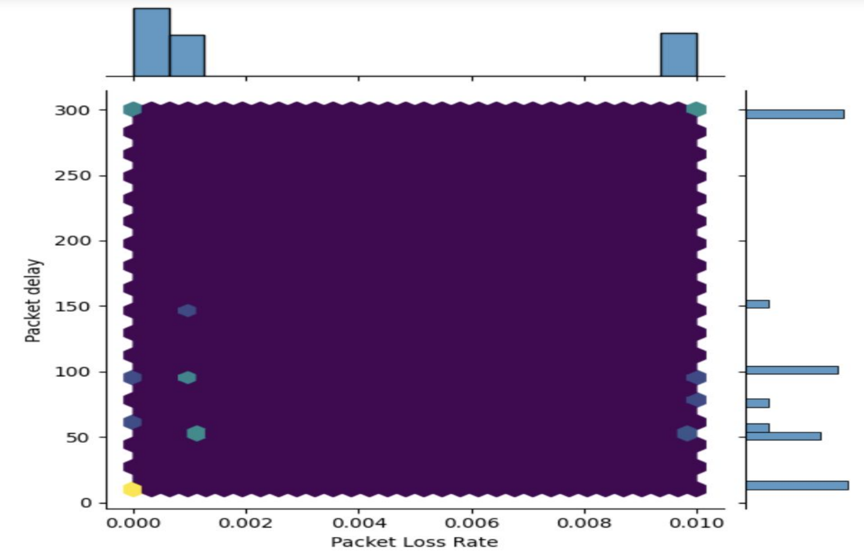
- The histograms provide a visual representation of the distribution and frequency of various LTE/5G network parameters.
- The 'Time' histogram appears to have a uniform distribution with all bars at similar heights.

Exploratory Data Analysis

Box plot of all input features



Join plot of two input features

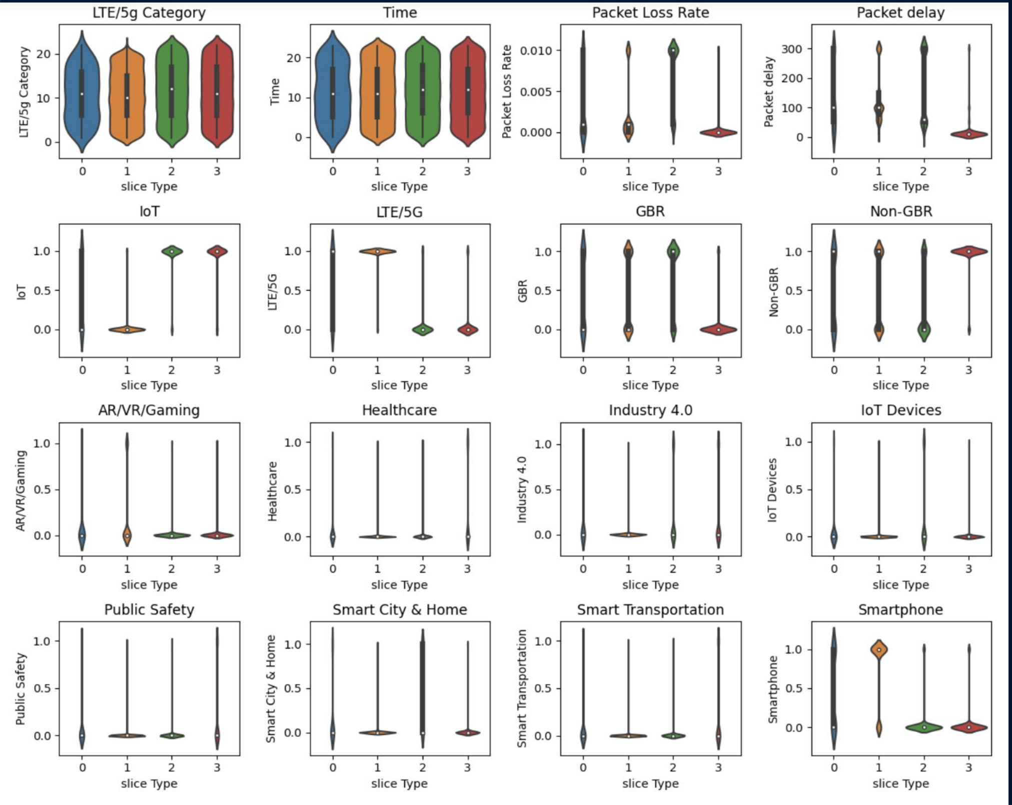


- The box plot illustrates the range, median, and quartiles for each LTE/5G network parameter, with the possible presence of outliers.
- This visualization implies that 'Packet delay' have a wide variation in values,while others have more concentrated range.

- In the join plot, each point represents an observation with its corresponding packet delay on the y-axis and packet loss rate on the x-axis.
- The histograms along the top and right margins of the join plot show the frequency distribution of the packet loss rate and packet delay, respectively

Exploratory Data Analysis

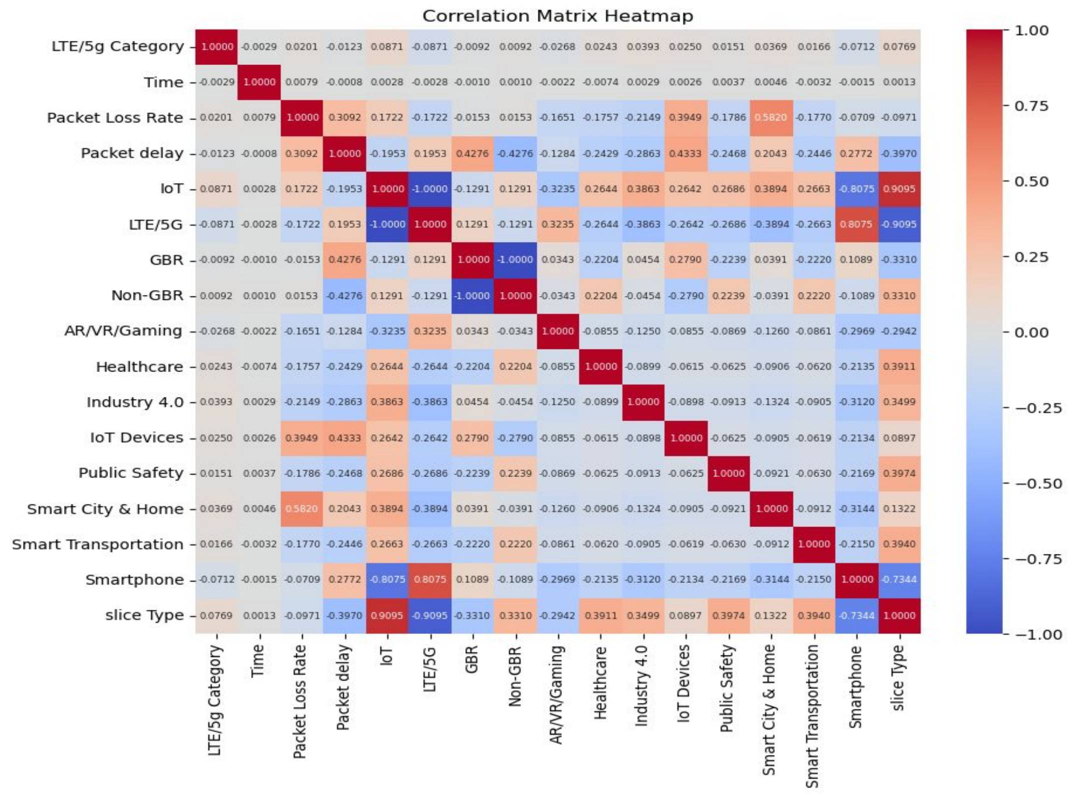
Violin plot of all input features



- violin plot provides insights into the distribution and range of values for the different slice types across various performance metrics and applications.
- From the above visualization we can observe that LTE/5G, Time are uniformly distributed.

Exploratory Data Analysis

Correlation matrix as a heatmap



- heatmap represents a correlation matrix of LTE/5G network parameters and device types are related to each other
- Highest correlation exist between Iot and slice type(0.9095)
- Lowest correlation exist between smartphone and time(-0.0015)

Data Preprocessing

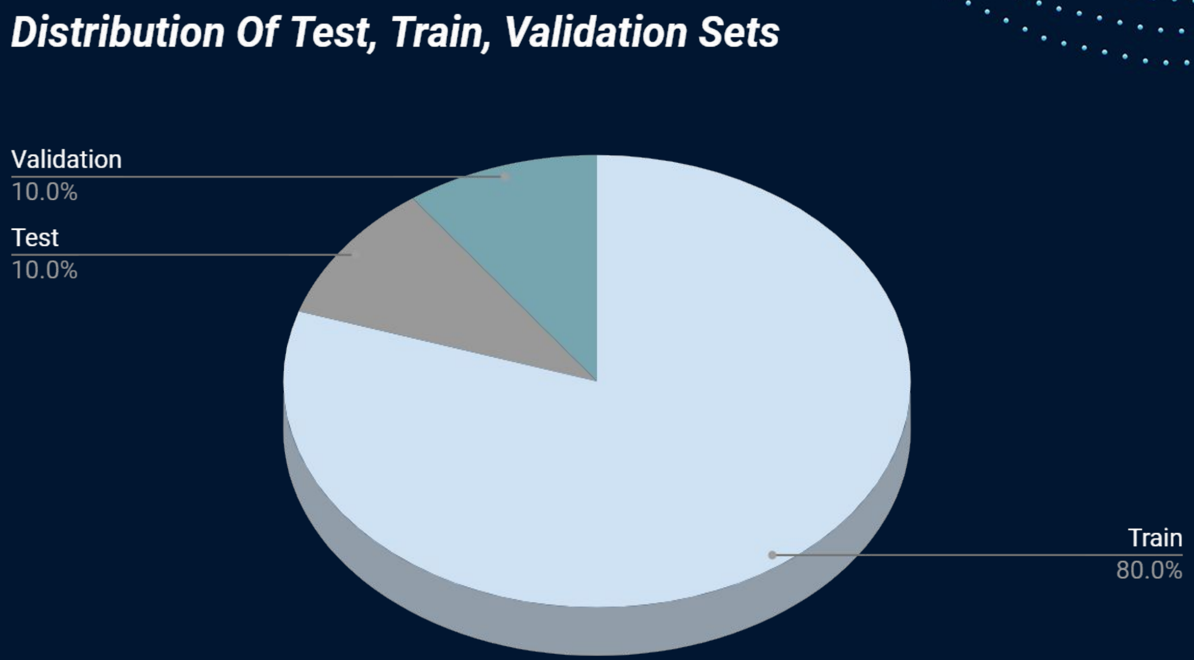
Data Normalization

	LTE/5g Category	Time	Packet Loss Rate	Packet delay	IoT	LTE/5G	GBR	Non-GBR	AR/VR/Gaming	Healthcare	Industry 4.0	IoT Devices	Public Safety	Smart City & Home
0	0.619048	0.000000	0.00000	0.000000	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0
1	0.809524	0.869565	0.09991	0.310345	0.0	1.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
2	0.761905	0.608696	0.00000	1.000000	0.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.095238	0.739130	1.00000	0.310345	0.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.380952	0.173913	1.00000	0.137931	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0

Highly Correlated Features

```
Highly correlated features with slice Type :
{'Smartphone', 'Healthcare', 'Smart Transportation', 'Public Safety', 'LTE/5G', 'Packet delay', 'IoT'}
```

Data Preparation



Machine Learning Model

- By utilizing preprocessed data, dividing dataset in to 3 sets train,test and validation with 80:10:10 ratio.
- Below are the steps that are followed in model pipeline
  - Base Models (SVM, Random Forest, Logistic Regression, Decision Tree, KNN)
  - AutoML with hyper parametric tuning
  - Ensembling Technique
- In the prediction pipeline we will be using the best model that we got at the end for prediction and validate the model using validation dataset.
- We are using F1 score and accuracy as our evaluation metrics for comparison as this is a classification problem.

Base Models

- We used the classifier models SVM, Random Forest, Logistic Regression, Decision Tree, KNN from keras as our base model.
- So we planned to use their evaluation scores in future experiments with hyper parameter tuning and ensembling.

Algorithm	Accuracy	F1 Score
SVM	0.887	0.870
Random Forest	0.915	0.902
Decision Tree	0.914	0.900
K-Nearest Neighbors	0.824	0.798
Logistic Regression	0.864	0.843

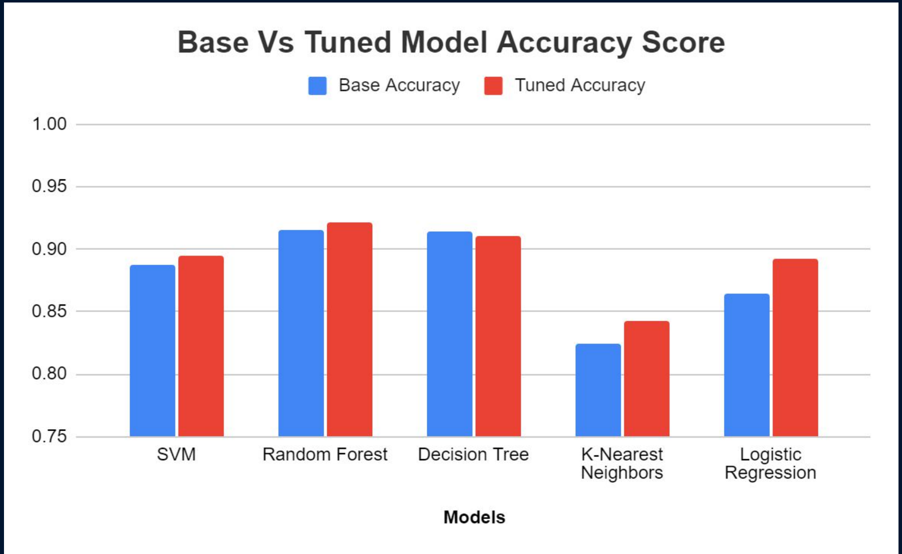
Auto ML with hyper parameter tuning

- In this pipe line we are using the base models, applying the hyper parameter tuning technique and selecting the best models for ensembling.

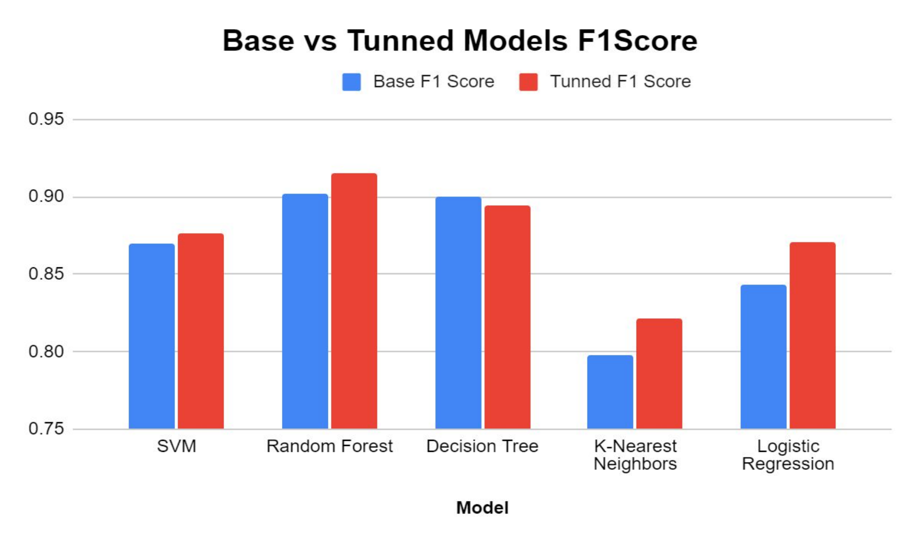
Algorithm	Best Parameters	Accuracy	F1 Score
SVM	**{'C': 0.1, 'kernel': 'linear'}**	**0.895**	**0.876**
Random Forest	**{'max_depth': None, 'n_estimators': 10}**	**0.922**	**0.915**
Decision Tree	**{'max_depth': None}**	**0.911**	**0.894**
K-Nearest Neighbors	{'n_neighbors': 5, 'weights': 'uniform'}	0.843	0.821
Logistic Regression	{'C': 0.1}	0.892	0.871

Model Evaluation

Accuracy Comparison

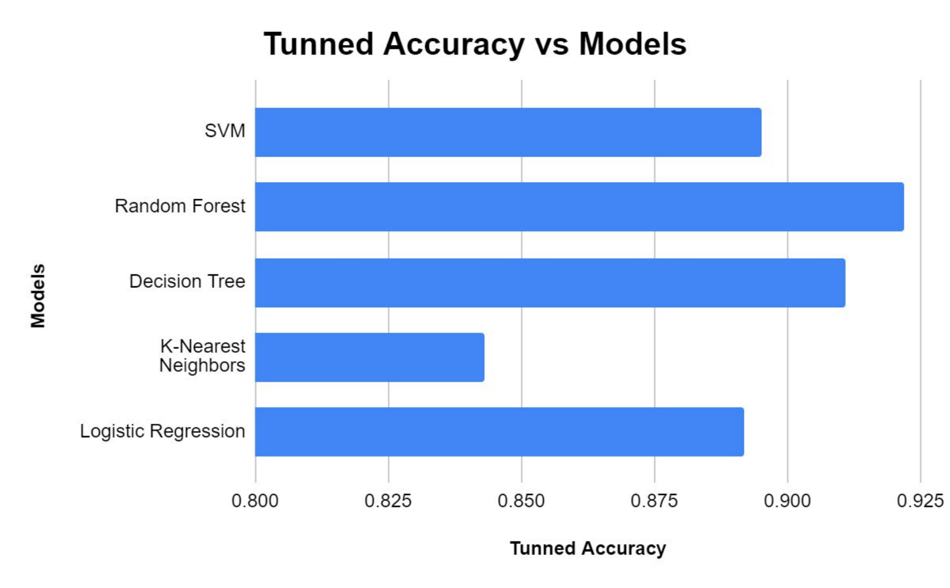


F1 Score Comparison

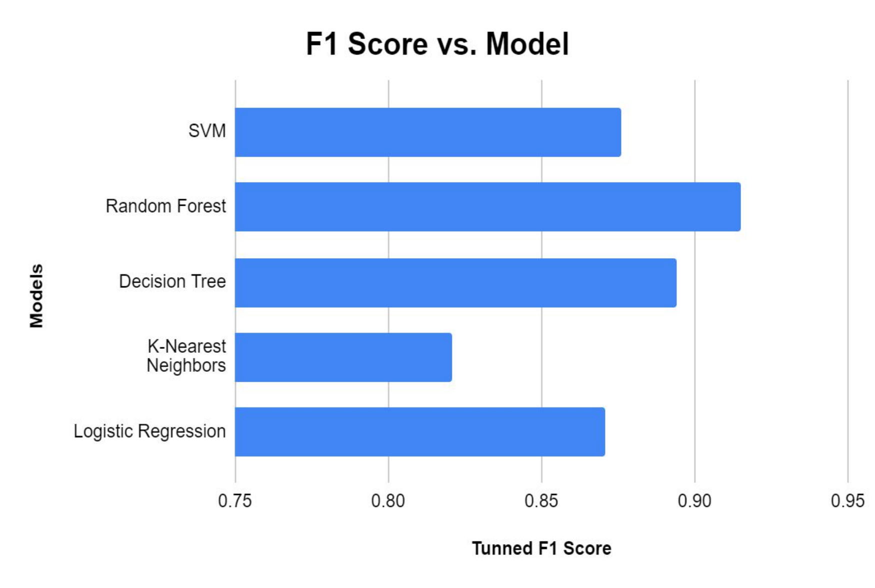


Model Comparison

Accuracy



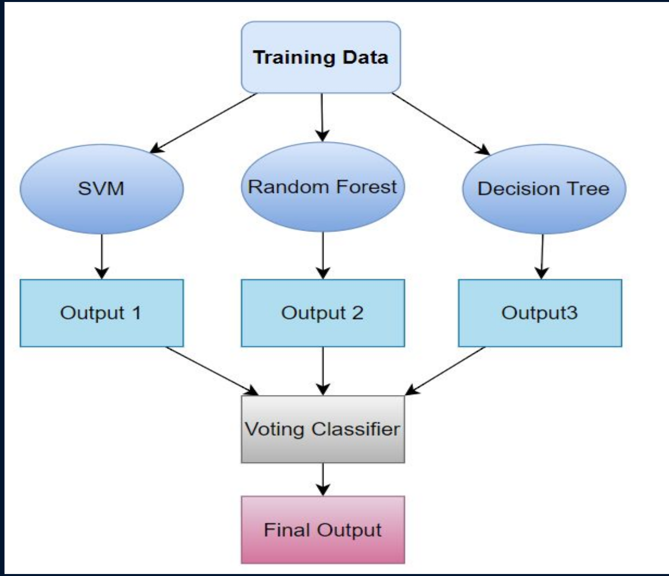
F1 Score



Ensembling

- As all the models are performing good so we are using the voting classifier technique to form an ensemble model.
- This Ensembling technique achieved the best accuracy of 95% out of all the models

Ensemble Model



Prediction Using the Best Model

Input

```
X_test.head(1)
```

	Packet delay	IoT	LTE/5G	GBR	Non-GBR	AR/VR/Gaming	Healthcare	Industry 4.0	Public Safety	Smart Transportation	Smartphone
21364	0.482759	0.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0

Predicted Output

```
grid.predict(X_test.head(1))
```

```
array([1], dtype=int64)
```

Actual Output

```
y_test.head(1)
```

	slice Type
21364	1

Model Deployment

SLICE TYPE PREDICTION

Packet Delay (The time taken for a packet to be received):	1
IoT (Device Usage):	-1
LTE/5G (User Equipment categories or classes to define the performance specifications):	-2
GBR (Guaranteed Bit Rate, Ensures a minimum rate of data transfer):	2
Non_GBR (Non-GBR services do not guarantee a specific data transfer rate):	0
AR_VR_Gaming (Augmented Reality (AR), Virtual Reality (VR), and Gaming applications):	-3
Healthcare (Usage in Healthcare):	2
Industry_4.0 (Usage in Digital Enterprises):	-2
Public Safety (Usage for public welfare and safety purposes):	3
Smart Transportation (Usage in public transportation):	-3
Smartphone (Whether used for smartphone cellular data):	1

Predict

Prediction: 3

Conclusion

- Through the integration of advanced machine learning algorithms and ensemble methods, we have developed a system capable of accurately predicting and optimizing network slices
- Our findings demonstrate that the use of AutoML not only simplifies the process of model selection and tuning but also significantly improves the performance and reliability of network slice classification.
- The ensemble technique employed further augmented the predictive power of the system, ensuring robustness against varied network scenarios and volatility.

Future Scope

- Exploring adaptive algorithms to dynamically adjust slice optimization in real-time network conditions.
- Extending the system to accommodate evolving network architectures and emerging technologies.
- Investigating anomaly detection mechanisms within slices for enhanced fault tolerance and security.

Reference

- Xenofon Foukas et al. “Network slicing in 5G: Survey and challenges”. In: IEEE communications magazine 55.5 (2017), pp. 94–100
- Shunliang Zhang. “An overview of network slicing for 5G”. In: IEEE Wireless Communications 26.3 (2019), pp. 111–117
- Muhammad Rehan Raza et al. “Machine learning methods for slice admission in 5g networks”. In: 2019 24th OptoElectronics and Communications Conference (OECC) and 2019 International Conference on Photonics in Switching and Computing (PSC). IEEE. 2019, pp. 1–3.
- Spyridon Vassilaras et al. “The algorithmic aspects of network slicing”. In: IEEE Communications Magazine 55.8 (2017), pp. 112–119.
- Git hub: https://github.com/Priyankaakula/Network-Slicing-Recognition/tree/main

Thank You