TELEWIRE
ANALYTICS

TRANSFORMING DATA INTO INSIGHTS FOR A BETTER TOMORROW

CONTENT

OUR TEAM

Kamaldeep Dhami

Data Scientist

Nimmo Usman

Data Analyst

Kirandeep Kaur

Data Engineer

ABOUT US

- Telewire Analytics is an analytics company which provides end-to-end data solutions.
- It collects and organizes, transforms and analyzes data from different domains.
- The analytics company is well known for its work on the optimization of network and compute resources leveraging data science solutions.
- There are 100 plus employees in the company and multiple departments.
- For every project, the project team is created where employees from multiple departments work as a team to achieve the project goal

OUR CLIENT

Telewire has a project to analyze network activity logs provided by **CONNECTWORK.** The logs contain metrics that are needed for data analysis, and the project involves extracting, storing, transforming, and building a model to detect anomalies in the behavior.

This project is an example of IoT analytics, which involves applying data analysis tools and procedures to the vast amounts of data generated by connected IoT devices. In this case, the data is coming from **sensors on cell towers.**

The goal of the project is to **optimize** the energy and resources consumed by the cell towers by detecting **abnormal behavior** in the network and adjusting the network configuration accordingly.

The project involves collecting, cleaning, transforming, and analyzing data from **logs** provided by Connetwork & The findings are expected to decrease the cost of running thousands of cell towers across the continent.

TOOLS & TECHNOLOGIES USED

Azure Cloud

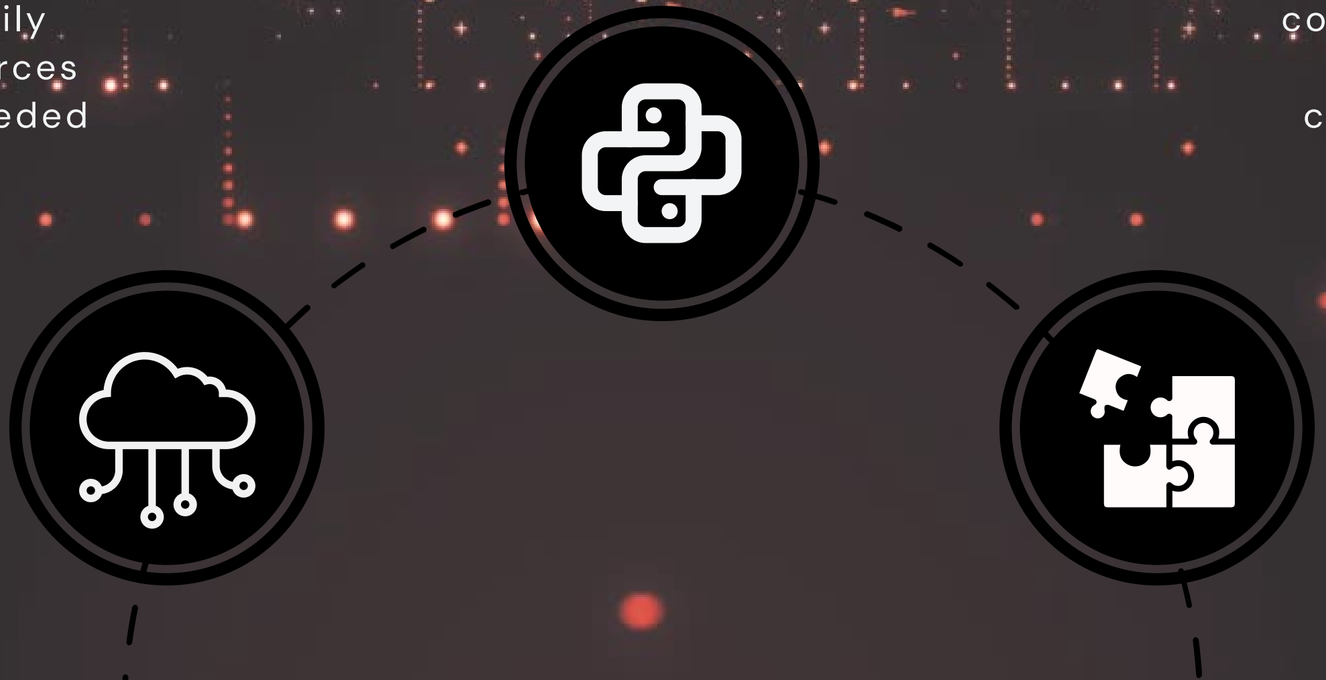
Azure cloud is a flexible and scalable infrastructure that allows organizations to quickly and easily scale their resources up or down as needed

Python

We used python because of its simplicity, versatility, large community, vast collection of third-party libraries

Power BI

Power BI is a powerful tool for data visualization, data integration, collaboration, mobile accessibility, customization, and security.



CELL TOWER ANOMALY

Cell tower anomaly refers to an abnormal behavior or performance of a cellular network tower, which may result in disrupted communication services for mobile devices. An anomaly can occur due to various factors such as equipment malfunction, software glitches, environmental changes, and network congestion.

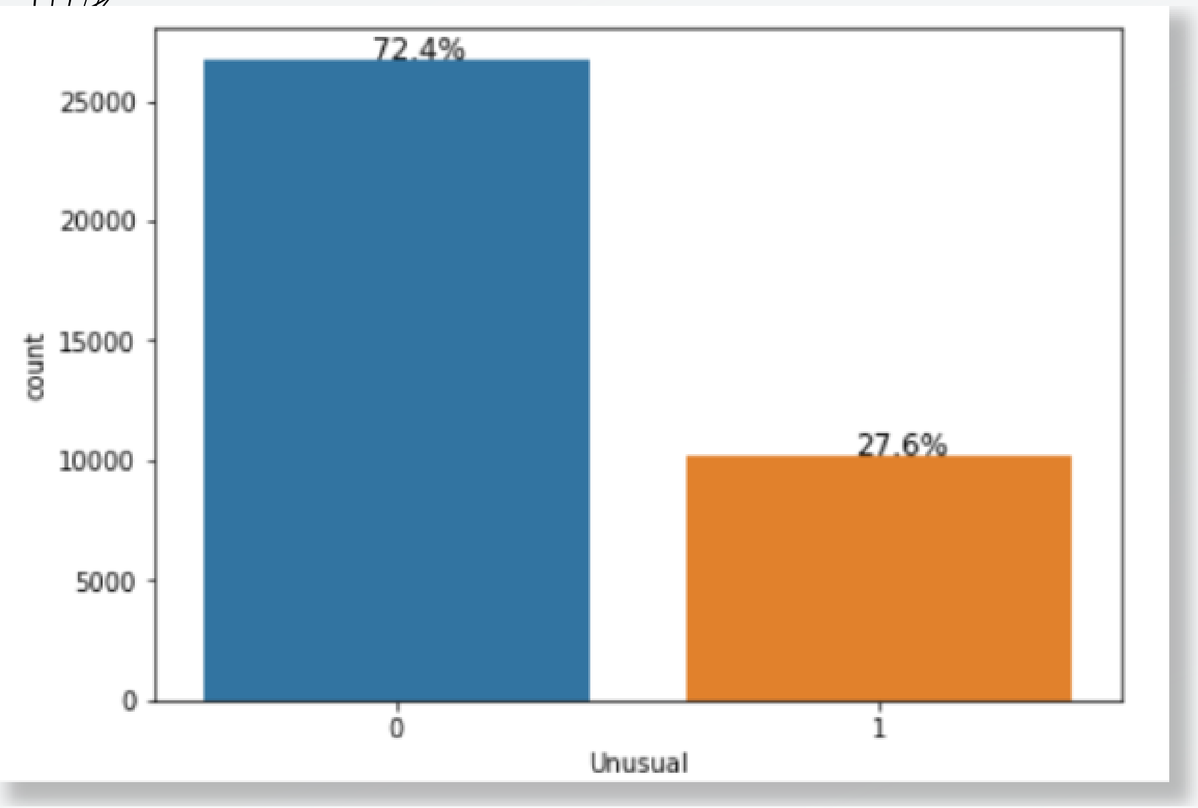
Some examples of cell tower anomalies include dropped calls, poor call quality, slow data transfer rates, and no signal or coverage. These anomalies can be caused by various factors such as equipment failure, interference, software bugs, and capacity overload.

ABOUT DATASET

The dataset has been obtained from a real LTE deployment. During two weeks, different metrics were gathered from a set of 10 base stations, each having a different number of cells, every 15 minutes. The dataset is provided in the form of a csv file, where each row corresponds to a sample obtained from one particular cell at a certain time

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Time	CellName	PRBUsage	PRBUsage	meanThr_	meanThr_	maxThr_D	maxThr_U	meanUE_[	meanUE_[	maxUE_Dl	maxUE_Ul	maxUE_Ul	Unusual
2	10:45	3BLTE	11.642	1.393	0.37	0.041	15.655	0.644	1.114	1.025	4	3	7	1
3	9:45	1BLTE	21.791	1.891	0.537	0.268	10.273	1.154	1.353	1.085	6	4	10	1
4	7:45	9BLTE	0.498	0.398	0.015	0.01	0.262	0.164	0.995	0.995	1	1	2	1
5	2:45	4ALTE	1.891	1.095	0.94	0.024	60.715	0.825	1.035	0.995	2	2	4	1
6	3:30	10BLTE	0.303	0.404	0.016	0.013	0.348	0.168	1.011	1.011	2	1	3	0
7	13:30	9ALTE	15.966	1.819	0.415	0.071	10.116	0.706	1.364	1.314	6	5	11	0
8	20:00	9BLTE	7.074	0.505	0.032	0.012	1.68	0.131	1.041	1.041	3	3	6	0
9	8:00	4BLTE	7.96	1.393	0.299	0.025	24.697	0.451	1.075	1.015	3	3	6	1
10	16:45	4CLTE	26.879	3.032	0.525	0.12	9.145	0.894	1.425	1.273	6	5	11	0

Preliminary analysis



- Initial examination and exploration of a dataset suggested we have 27.6% Unusual cases in the Dataset which are anomalies
- These anomalous instances could represent faults, errors.
- They need to be further investigated to determine the cause and take corrective actions

Checking For Missing Values

	Total Missings	Missing Count %	Total Missings (Unusual)	Missing Count (Unusual) %
maxUE_DL	89	0.2	16	0.87
maxUE_UL	89	0.2	16	0.87

Missing data in the dataset and the proportion of missing values is very small, only about 0.3% of the total data. However, among the missing data, there are 2 cases that are considered "Unusual," which means they are anomalies or outliers. These two cases represent about 1% of all the unusual cases in the dataset. Considering that 27.6% of total dataset are Unusual cases, it makes sense to drop the rows with missing values.

Heatmap to Visualize Correlation Between Feature Variables.

The features maxUEDL & maxUEUL are highly correlated based on the heatmap.

This means that there is a strong relationship between these two features, and changes in one of these features are likely to be reflected in changes in the other features as well.

Model Evaluation

Model evaluation is a critical step in the machine learning pipeline, as it helps to ensure that the model is performing well on the given dataset, and can be used to make accurate predictions on new data.

	precision	recall	f1-score	support
0	0.85	0.99	0.92	7995
1	0.97	0.55	0.70	3050
accuracy			0.87	11045
macro avg	0.91	0.77	0.81	11045
weighted avg	0.89	0.87	0.86	11045

Accuracy of random forest: 0.8709823449524672

	precision	recall	f1-score	support
0	0.72	1.00	0.84	7995
1	0.00	0.00	0.00	3050
accuracy			0.72	11045
macro avg	0.36	0.50	0.42	11045
weighted avg	0.52	0.72	0.61	11045

Accuracy of SVM: 0.7238569488456315

	precision	recall	f1-score	support
0	0.98	0.98	0.98	7995
1	0.95	0.94	0.95	3050
accuracy			0.97	11045
macro avg	0.97	0.96	0.96	11045
weighted avg	0.97	0.97	0.97	11045

Accuracy of decision tree: 0.9712086917157084

Model Evaluation

XGBoost:

	precision	recall	f1-score	support
0	1.00	0.98	0.99	8121
1	0.96	1.00	0.98	2924
accuracy			0.99	11045
macro avg	0.98	0.99	0.98	11045
weighted avg	0.99	0.99	0.99	11045

XGBoost Accuracy: 0.9866002716161159

XGBoost:

	precision	recall	f1-score	support
0	1.00	0.95	0.98	8383
1	0.87	1.00	0.93	2662
accuracy			0.96	11045
macro avg	0.94	0.98	0.95	11045
weighted avg	0.97	0.96	0.97	11045

Tuned XGBoost Accuracy: 0.9645088275237664

We can observe that XGBoost outperformed other models in terms of accuracy, precision, recall, and F1-score for both classes.

Neural Network

```
Epoch 200/200
806/806 [==============================] - 0s 533us/step - loss: 0.5878 - accuracy: 0.7246

Out[49]: <keras.callbacks.History at 0x1e3fe582fd0>
```

```
In [50]: # Train and Test accuracy
scores = model.evaluate(X_train,y_train)
print("Training Accuracy: %.2f%%\n" % (scores[1]*100))
scores = model.evaluate(X_test,y_test)
print("Testing Accuracy: %.2f%%\n" % (scores[1]*100))
```

```
806/806 [==============================] - 0s 409us/step - loss: 0.5875 - accuracy: 0.7245
Training Accuracy: 72.45%

346/346 [==============================] - 0s 647us/step - loss: 0.6078 - accuracy: 0.7231
Testing Accuracy: 72.31%
```

The First line indicates that the model was trained on 806 samples and achieved a training accuracy of 72.45%. The loss function used during training was 0.5875

The second one indicates that the model was evaluated on 346 samples from the testing dataset and achieved a testing accuracy of 72.31%. The loss function used during evaluation was 0.6078.

CONCLUSION

Based on the analysis of the various models, it can be concluded that the Random Forest and Decision Tree models perform the best with accuracy scores of 87.10% and 97.17% respectively. The XGBoost model also performs well with an accuracy score of 98.66%, but the tuned version of the model has a slightly lower accuracy score of 96.45%.

On the other hand, the SVM model has a much lower accuracy score of 72.39%, and the neural network model has a similar accuracy score of 72.31%. Therefore, it can be concluded that the SVM and neural network models do not perform as well as the other models in this specific task.

LET'S STAY IN TOUCH

Do not hesitate to ask us further about the project here:

+123 456 7890

info.telewire@gmail.com

TRANSFORMING DATA INTO INSIGHTS FOR A BETTER TOMORROW