Capstone Project Proposal

Bertrand Ndakena

Business Goal

Project Overview and Goal

What is the industry problem you are trying to solve? Why use ML/AI in solving this task? Be as specific as you can when describing how ML/AI can provide value. For example, if you're labeling images, how will this help the business?

What problem am I trying to resolve?

I will start by defining who a **NOC** (Network Operations Center) engineer is. NOC Engineers monitor large computer networks and servers for problems from a central location.

- They analyze problems

- perform troubleshooting and incident response on the system

- communicate with site technicians and

- track problems through to resolution.

The employment market for NOC Engineers is strong, as the telecommunications and IT industries continue to grow and become more complex. The Bureau of Labor Statistics predicts a job growth of 12 percent by 2024 for all computer support specialists, including NOC Engineers.

I happen to be working with Huawei Cameroon, where we do monitoring of MTN's entire network, including the backbone network. We actually monitor network terminal equipment like, BTS(Base transceive Station) for **GSM** or **2G** networks, Node B for **3G** network and eNodeB for **4G** or **LTE**(Long Term Evolution). A network **site** has **2G**, **3G** and **4G**. We face major issues in the process of detecting network problems (2G,3G and or 4G) and going through all the steps as stated above.

When a network site goes down, it sends alarms to the NOC engineer and the NOC engineer, after detecting the issue,

> Makes first-hand troubleshooting by checking the reported alarms to determine which team will handle the issue (though other troubleshooting actions are carried out; this is the most important action required to be carried out by the NOC engineer. Other actions are required to be done by appropriate teams).

The NOC engineer creates an incident ticket and assigns it to the appropriate team (I.H.S, FME or BO) as per the reported alarms.
He then goes ahead to notify the team by a phone call, SMS, and or email
Does the follow up until the site restores?

Now, the problems here are;

i. **Long Downtime:** It takes a relatively very long time for a NOC engineer to complete this process each time a site goes down, thereby increasing the "downtime". A longer site's downtime implies bad customer experience and reduction in revenue.

ii. **Incidents go unnoticed:** Some incidents(like site fluctuation) go unnoticed because the downtime is usually very short and restores before the NOC engineer notices. When incidents go unnoticed, it implies no incident ticket will be created. Incident tickets are meant to account for site's downtime, indicating that the site was unavailable during the period. KPI calculations take into consideration the downtime as per the trouble ticket. If we do not have tickets to account for the downtime, the KPI will result in a false value as it will consider that the network was available throughout.

** By saying the **site is down**, we imply that it is unavailable and mobile users within it cell are out of service.

Now how do I use ML to resolve the issue of **Long Downtime** *and* **unnoticed incidents***?*

To resolve these two main issues, I will like to build an AI system that will

1. detect an incident(site down)
2. collects all reported alarms
3. Use a trained model to determine the appropriate team that should resolve the issue. This will be by simply flagging the site's ID with either (IHS, BO or FME). This is a 3-class classification problem.
4. create an incident ticket and assign it to the team concerned
5. send an SMS notification to the team concerned about the incident.

How will this benefit the business?

Reducing incident's downtime and number of unnoticed site issues implies;

* High network availability(the company sells network availability to gain revenue)
* Better customer's experience
* Increase in revenue.

Business Case Why is this an important problem to solve? Make a case for building this product in terms of its impact on recurring revenue, market share, customer happiness and/or other drivers of business success.	• The revenue of any telecom company depends on the availability of its network. If the network is mostly available, customers will be happy to use the services of that network operator and there will be low customer churn. More customers will be attracted too, hence high market share and increase revenue. • Resolving this problem will also mean better and faster incident management with less number of NOC engineers.
Application of ML/AI What precise task will you use ML/AI to accomplish? What business outcome or objective will you achieve?	I will use ML to classify site problems into 1. **IHS** problem: IHS is a 3rd party company that manages the power supply for MTN Cameroon, any problem with power on site will cause the site to go down and should be managed by IHS. 2. **BO** (Back Office) problem. If a site is down and there is no power issue on site, depending on the reported alarms, the issue can be managed by the BO engineers. 3. **FME** problem: If BO engineers cannot handle the issue, the FME (Field Maintenance Engineers) will need to intervene on site. Also, we could assign an incident to the FME directly and all depends on the alarm generated as the site goes down. I intend to increase the company's revenue by providing better and faster incident management. Poor and slow incident management reduces network availability.

Success Metrics

Success Metrics What business metrics will you apply to determine the success of your product? Good metrics are clearly defined and easily measurable. Specify how you will establish a baseline value to provide a point of comparison.	In the company, we have a metric call network percentage availability, if no site goes down for the who day/week/month, we will have an availability of 100%. Now since it is impossible to go for a day without having an issue with at least some few sites, the availability is always below 100%. 1. If our product works fine, it will have an immediate impact on the availability(which is an outcome). A 24hours availability (A24) of a site is given by; **Daily Availability**: $A_{24} = \frac{24-Dt}{24} x100\%$, Dt = downtime in hours **Hourly Availability**: $A_{60} = \frac{60-Dt}{60} x100\%$, Dt = downtime in minute 2. While building our system, there are other metrics we will use to attempt to measure the success of our ML model. As much as the most important thing is the availability, if our model fails us, the system will fail us too. Also, we must evaluate the success of our model, using some modeling metrics so that we will be able to identify if there are biases or not and to be able to improve on the whole system. With that said, we will use the metric that takes into

	consideration both **recall** and **precision** as they are very important, the **F1 score**. 3. Regarding the baseline for comparison, we will use **A/B testing**. Since we have a working system(**Control-A**) and the new AI system(**Treatment-B**). We will divide the number of sites between these two systems(**80% for A and 20% for B**) and their availability values can be compared.

Data

Data Acquisition Where will you source your data from? What is the cost to acquire these data? Are there any personally identifying information (PII) or data sensitivity issues you will need to overcome? Will data become available on an ongoing basis, or will you acquire a large batch of data that will need to be refreshed?	- The source of my data to be used to train the model will be the current working system. We will simply need to extract past incidents(which were either assigned to IHS, BO or FME). - There will be no cost to get the data as the data is accessible by NOC engineers and they are free to use the data to improve on incidents management, especially for recurrent incidents. - There are no sensitivty issues the data to be used for the model requires only site **IDs**(which are simply numbers), the assigned group**(IHS, BO or FME**) and the reported alarms(**Mains failure, generator failure, low battery, low fuel, DC power abnormal, Undervoltage, ethernet down, device offline, RRU disconnected, digital cable failure, lost communication to radio unit, link failure, low input power, license expired, ...**). When a site goes down, it sends one or many of the alarms above to help determine the problem and the appropriate team. - Data will always become available on an ongoing basis. For unknown cases, the ML learning model can assign the tick to the few NOC engineers. After troubleshooting, NOC engineer will reassign to the appropriate team, the ML will then pick this particular case and train itself on(Continuous learning).
Data Source Consider the size and source of your data; what biases are built into the data and how might the data be improved?	- Regarding the data size, there is a large data size (millions of samples), however, there are actually biases as there may be incidents tickets that were assigned to a wrong group before closure. This will bring about the idea of **false positive(FP)** and **false negative (FN).** - To do away with this issue, we may have to collect new data instead of just extracting for the old incidents. In the collection process, we make sure that eliminate these biases and that all power problems are assigned to IHS, all back office problems are assigned to BO and all problems managed on-site assigned to the FME.
Choice of Data Labels What labels did you decide to add to your data? And why did you decide on these labels versus any other option?	As we have three main teams involved in incident management, I used the name of those teams as labels. as already mentioned above, we depend on alarms to label sites or incidents as **IHS, BO or FME. The labels will be as below, it shows which label to select for each corresp[onding alarms.**

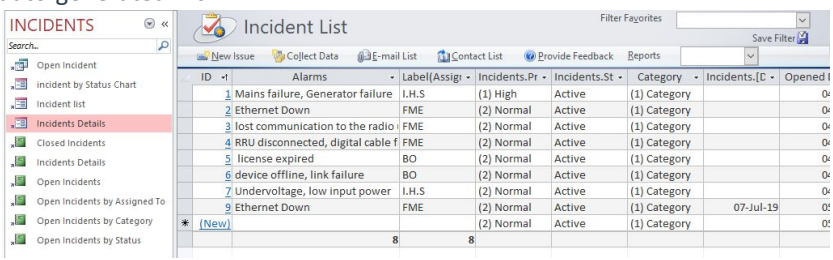
IHS	BO	FME
Mains failure, generator failure, low battery, low fuel, DC	**license expired,**	**ethernet down, RRU disconnected, digital cable failure, lost**

	power abnormal, Undervoltage, low input power,	**device offline, link failure, sector failure...**	**communication to the radio unit...**

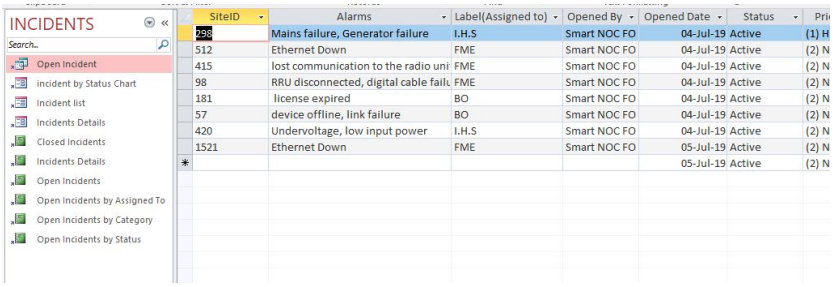
Model

Model Building How will you resource building the model that you need? Will you outsource model training and/or hosting to an external platform, or will you build the model using an in-house team, and why?	I will consider using a hosted platform to build and train the model as this eliminates the need for hardware in-house such as servers. Instead, the vendor(Google), will provide the latest and greatest server technology. This is a good idea as Google upgrades their server resources more frequently. Unexpected hardware purchases can be very costly, so this can be a very huge cost saving.
Evaluating Results Which model performance metrics are appropriate to measure the success of your model? What level of performance is required?	Let me start by defining what precision and recall are; - Precision is the measure of total relevant results obtained by the model - Recall is the measure of total relevant results correctly classified by the model. Now, we note that if we are trying to obtain a very high precision(100%), this must be at the expense of the recall and vice versa. Since we need to recall as much as we can and at the same time be more precise, it is important to use a metric that takes into consideration both precision and recall. This implies that we are going to use **F1 score** which is simply the harmonic mean of precision and recall. $F_{1-score} = 2x\frac{Precision\ x\ Recall}{Precision + Recall}$ Since the model is not aimed at eliminating the functions of NOC Front Office engineers, 60% performance is great.

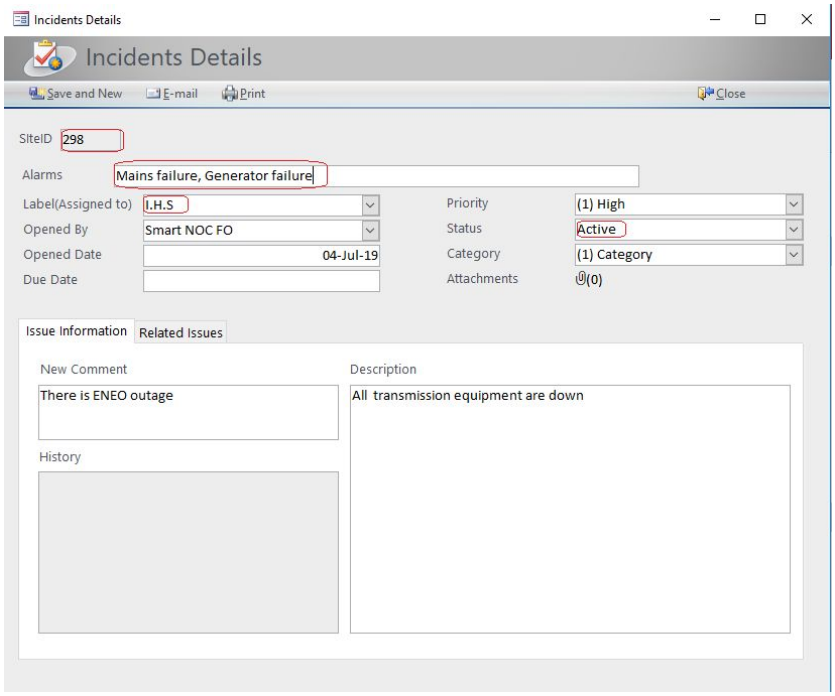
Minimum Viable Product (MVP)

Design What does your minimum viable product look like? Include sketches of your product.	The MVP will look like below. please check the "screenshots" folder in my submission to see the image clearly. This image displays "Incident list" with auto-generated IDs. 

The nex image displays open incidents, including the “SiteID” and the “assigned to” field. The Assigned to is equivalent to the label selected by the **model**. “The SiteID” field represents the site that is actually down. The “Alarms” field represents the Alarms that the model considers to assign the Incident to the appropriate group.



There are other functions in the product that are not necessary to explain, but I have included more screenshots.



Use Cases

What persona are you designing for? Can you describe the major epic-level use cases your product addresses? How will users access this product?

I am designing for telecom operators. IT will be used by network monitoring engineers. The major epic use cases of my products are:

1) detects incident (site down)
2) Collects reported alarms.
3) Label the site with either IHS, BO or FME according to the alarms reported
4) Creates and assigns an incident ticket to the group with the same name as the Label name.

The product will be installed in a central server at NOC for all Engineers to be able to access.

Roll-out

How will this be adopted? What does the go-to-market plan look like?

This product will be used at NOC and customers won’t need to access it. SO, we will simply need to train engineers on how to use it after the A/B testing favors the new AI product.

Post-MVP-Deployment

Designing for Longevity How might you improve your product in the long-term? How might real-world data be different from the training data? How will your product learn from new data? How might you employ A/B testing to improve your product?	TO improve on my product, - I will make the model is learning continuously for new cases(new alarms) - Consider feedback from users of the product - Monitor the outcome
Monitor Bias How do you plan to monitor or mitigate unwanted bias in your model?	I intend asking users of the product to report any case of wrongly assigned incidents, so that I will be able to track where it is coming from. I could also include another label for cases where a site goes down with no alarms, this can be assigned to the NOC FO engineer for him to manually determine the appropriate group, feed the case into the model for it to keep learning. An abnormal high availability could also depict a bias in the data.