**Tell me about yourself.**

Hello Eri, my name is Zheng Yang. I have graduated from NTU Mechanical Engineering with a First class honour in 2019 and will be pursuing my Msc in Industry Engineering from this Aug onwards. My current role is contact engineer and data analyst for my department. My role as a mechanical contact engineer is to protect the mechanical integrity of equipment by conducting failure analysis and ultimately enhancing the reliability and sustainability of my plants by eliminating bad actors.

As an analyst, I consistently worked with huge amount of data where I analyzed data systems, automating information retrieval and preparing reports to ensure tasks, costs compliance and laws compliance on equipment like pressure vessel to ensure we did perform inspection on register few K pressure vessel in our site as per MOM law. Furthermore, I also have led a team to deliver business intelligence which has helped my department to save up to 50k SGD per month by identifying barriers and spearheaded machine learning project on prediction of remaining thickness of piping which is susceptible to corrosion.

Even thought I am relatively new to data science field, I quickly found out this is where my passions lie, and I have spent a lot of free time to learn and develop the relevant skillsets which Data scientists should acquire. For instance, I went for research and initiated machine learning project even though I am occupied with my contacting role in business unit. Moreover, to further build up my portfolio in DS realm, I went for Tableau certificate and AWS cloud exam and passed them with flying colours.

**Why Micron/Data Science?**

I always wanted to work in an industry which has a vast amount of growing opportunities. Before covid19, there was a huge amount of growing opportunities in the OnG industry, but things have changed drastically. Covid19 has disrupted businesses, a lot of traditional brick and mortars store had laid off a lot of employees across multiple countries, people having multiple lockdowns from time to time or forced to work from home. The amount of time people spend online has gone up tremendously, the sudden surge in remote work just accelerates digital infrastructure adoption.

Most of the consumers turned to online retailers which has given tech and semicon industries a great boost. Coronavirus has accelerated process of digitalization. With the rapid growth in Tech and e-commerce sector, I really want to a part of Micron’s data science team to transform how the world uses information to shape a better future for all.

Hence when I saw your job posting on LinkedIn, I did not hesitate to apply for it. I believe my technical, problem solving skills that I have picked up along the way and my courage to seek gap for delivering value-adding results to reach company goals.

**Why you choose to study Industrial Engineering**

Why mechanical engineering, because I want to solve real world problems using physics and maths. After working, I realize that I rather like looking processes and systems at a whole, not just one piece. From a overall system perspective, IE can allows me can adopt a systematic approach to improve, innovate and develop integrated real-world systems coupled with modern engineering and IT tools. Morevoer, it allows me to combine technical skills (mechanical engineering) with business acumen which I think it will make me to become a more valuable employee

But with IE, I can use solve real world problems using not only physics and math, but also computer science, economics and business.

1. It opens up a new career path for me (open up more doors)
2. It allows me to combine technical skills (mechanical engineering) with business acumen.
3. Analyze, Innovate, Optimize, Decide
4. I like working with

**What do you see yourself in short term/5 years**

In the short term, I hope to work as a Data science related role. As a tech-oriented person, I would like to work on my expertise on Machine Learning and relevant DS skillset to enhance my competency.

I wish I can grow together with a company where I can continue to learn, enhancing my skills and providing positive output for the company. Moreover, hopefully I am able take on expanded technical leadership responsibilities in the future, as these become available.

In 5 years’ time, tools that we are using currently might not be relevant, data science discipline might have a revolution and get more interesting. However, the impact of DS will still stay. That being said, in the next few years, I would like to keep up with the latest technologies and get better at solving problems at larger scales as a senior technical engineer or lead engineer.

**Technical Qs**

1. OLAP and OLTP

OLAP is online analytic processing, it is software tool for pulling data from multiple database for analysis purpose. Date warehouse is an example of OLAP. (data warehouse is large collection of business data used to help an organization make decisions). RWUD

OLTP is online transitional processing. It is used for operational purpose for transaction purpose. It is used for maintain the online transaction. (online booking system, read in millisecond)

1. JOIN

Join is a way to combine rows from 2 or more tables together. The type of joins contain inner join, left join, right join, full outer join. (if we didn’t mention, default join is INNER Join)

* (INNER) JOIN: Returns records that have matching values in both tables
* LEFT (OUTER) JOIN: Returns all records from the left table, and the matched records from the right table
* RIGHT (OUTER) JOIN: Returns all records from the right table, and the matched records from the left table
* FULL (OUTER) JOIN: Returns all records when there is a match in either left or right table

**Machine Learning Questions**

Start by explaining what’s algorithm about. What is the advantage & disadvantage?

KNN: Compute Euclidean distance between points, a classifier, good for small dataset, easier to explain

Naïve bayes: “Naïve” because the algorithm treats every features is independent to one another, not so accurate, good for large dataset,

Decision tree: Making multiple if else questions based on the scenario, tree node, decision node, leaf, easy to explain, no need to scale

Random forest: Ensembled method of decision tree, contain multiple decision tree where individual tree is slightly different than another one which reduce overfitting by averaging the result since multiple decision trees can cause overfitting. Don’t need to scale data like decision tree, very robust and powerful

SVM: Need to scale data and sensitive to hyperparameter. Kernel trick to promote to high D without needing too much computational resources for linear separation (hyperplane). SVC uses hyperlane to classify class by promoting the current linear inseparable features into higher dimensional.

Gradient boosted decision trees: Another ensembled method, instead of the having multiple decision trees. It build trees in a serial manner which each trees corrects the mistakes of the previous one. which has a slightly higher accuracy than random forest

**Micron Core Values**

Micron Technology is a **world leader in innovative memory solutions that transform how the world uses Information**. Through our global brands — Micron, Crucial and Ballistix — we offer the industry’s broadest portfolio, and are the only company that manufactures today’s major memory and storage technologies: DRAM, NAND, NOR, and 3D XPoint™ memory

Micron Technology is a world leader in innovating memory and storage solutions that accelerate the transformation of information into intelligence, inspiring the world to learn, communicate and advance faster than ever. We deliver the world's broadest portfolio of technologies at the core of today's most significant disruptive breakthroughs such as artificial intelligence and autonomous vehicles.

**Roles & Responsibilities**

As a Data Science Engineer at Micron Technology Inc., you will be a key member of a cross-functional team responsible for developing and growing Micron’s methods and systems for extracting new insight for our expanding data streams. You will be collaborating with other data scientists, solution analysts, engineers, technicians and data mining teams to design and implement systems to extract data from Micron’s business systems, transforming it into an actionable format, and as needed, creating dynamic presentation layers for use by high-level engineers and managers throughout the company. You will be creating new solutions, as well as, supporting, configuring, and enhancing existing solutions.

**Responsibilities and Tasks**

**Understand the Business Problem and the Relevant Data**

* Maintain an intimate understanding of company and department strategy
* Translate analysis requirements into data requirements
* Identify and understand the data sources that are relevant to the business problem
* Develop conceptual models that capture the relationships within the data
* Define the data-quality objectives for the solution
* Be a subject matter expert in data sources and reporting options

**Develop, Support, Automate, and Orchestrate an Ecosystem of ETL Processes for Varying Volumes of Data**

* Identify and select the optimum methods of access for each data source (real-time/streaming, delayed, static)
* Determine transformation requirements and develop processes to bring structured and unstructured data from the source to a new physical data model
* Develop processes to efficiently load the transform data into the data management system
* Troubleshoot operational escalations and resolve business partner issues in a timely manner with strong collaboration and care for business priorities.

**Generate Insights**

* Discovering new solutions and opportunities by analyzing data, identify patterns and trends
* Collecting & modelling large sets of structured and unstructured data from disparate sources
* Devising and applying models and algorithms for mining data.
* Working with unstructured data like video, images, etc

Job as an analyst:

We have around 40 inspectors and technicians and have around 1000 planned tasks and 500 unplanned tasks per year. With so many tasks in hand, we need a concrete plan to

1. Identify goals to be achieved
2. Capture the best metrics to reach goals
3. To catalog resources
4. Assign responsibilities for controlling and consuming inspection resources.

Talking about overseeing the overall health of your department and how other department perceive our department, there isn’t a better way to represent those without a business intelligence tool. In the beginning, we do track multiple KPIs, but our managers need to have multiple meetings with their respective superiors because the data is so discrete and scattered around.

I have championed opportunity to drive business intelligence application using Tableau to consolidate all the KPI’s excel spreadsheet to Tableau. (I am chemical plant), after setting up the template, I also guide other 2 sites in SG to Tableau and setting up path to my interactive Tableau template to feed in the data accordingly. The teammates where I have does not even have good excel skills let alone skill to do data analysis. I encouraged, very patient, meticulous when doing the coaching and they were able to complete their responsibilities as mine. We were able to complete this project successfully on deadline and which are able to transform data at hand into deeper insights for management to make better decisions.

**CUI IDX**

Background:

The most common damage mechanicsm of my equipments/piping are CUI. Piping are insulated to mainitain process temperature/act as a barrier to prevent water ingression. Current method is strip the insulation, and check the piping condition. Sometimes some of the pipings are high up in the sky, and very often we need to erect scaffold for workers to strip off the insulation and conduct visual inspection. The method is more accurate, but it will cost the company a lot of money for scaffolder erection and resources.

We used risk based inspection to determine the inspection frequency of equipment/piping in our database. Risk based inspection is the process of developing a scheme of inspection based on knowledge of PoF and consequence of failure. When we followed the risk based inspection, the hit rate (findings/# of time we open) is around 25-30%, which means the risk based inspection is way 2 conservative. Over few million SGD per year has been spent on those resources to open up the insulation and building scaffolds. After understanding the background , I realize that ML might be coming in handy to help to predict the remaining thickness of piping/equipment to allow us to relook at inspection frequency for CUI tasks. For RBI, instead of just numerical value, we classified the remaining thickness of respective equipment and piping into several categories, and we will have different strategies to mitigate the risk on respective thickness categories.

**Intent**: To predict the remaining thickness of equipment or piping those are going to leak as per risk based inspection criteria using given features.

**Data Source**: The data source comes from a team of engineers who build the database which contains the relevant information of CUI inspection tasks that we have done for the past 7 or 8 years. The database captures features like year of service, operating temperature, type of insulation, coating type, coating age, presence of heating element(steam tracing, electrical wire) original reading, corrosion rate and min. capture reading.

The data source wasn’t quite neat in the beginning as expectation then vs now is very different and amount of features we capture last time is less than now so data cleaning is inevitable. Some techniques that I have used to prepare the data are, one hot encoding to handle categorical data like type of insulation and coating. For corrosion rate, missing data are replaced with a corrosion rate of that service according to industrial documents. Numerical data are scaled as well to ensure all features are on the same scale. We are facing using of imbalanced dataset as I have said before, we are getting a low percent hit rate on CUI inspection. So around 60 percent of the time, we will find minimum thickness loss. So what I have done is to resample the training set by under sampling the majority class of data. I used a 80:20 train test split to ensure there is no data leakage.

**Performance metrics:**

Accuracy, recall is also important because we dun want to underestimate the consequence which means the predicted wall loss is lesser than actual wall loss. Hence, we want to boost the algorithm result on recall metric on all classes. F-score of multiclass version also to reduce both false positive and negative

**Baseline model**

After sorting out the cleanliness and performance metrics for the data. I proceeded with modelling and I started on working out on my baseline model which can be used for comparison with more complex models. I choose KNN model as my go to model since it is relatively easy to explain to non-technical people and it can provide a relatively good accuracy.

**Model Training**

I begin with basic spot-checking several algorithms like support vector machine, random forest, naïve bayes using cross-validations, followed by selecting the one with the highest value of performance metric which I mentioned just now. After the cross-validation, it always comes down to two or three algorithms which vary only slightly in their performance. For this case, my to-go is ensemble model like random first since they tend to increase prediction accuracy by combining the predictions from multiple models together.

**Model Tuning Process**

I then proceeded with hyperparameter tuning using GridSearchCV and was able to achieve an F1 score equal to 80%. I also tried to undersample the majority class but it didn’t help much in improving performance. What did help was doing some feature reduction and finally my F1 score was around 82%.

Going forward, since I have gone through past inspection data till 2020. Next stage will be running the algorithm over 2021 & 2022 site wide data. If that success, I will wrap up the model in a nice little container in a form of API (interface that let you access and manipulate data in the backend when a user requests the data from the front end).

If this is successful, I am hoping to reduce maintenance cost for CUI inspection by a great amount (up to few hundred’s K !)

**Tableau Project Experience**

Business intelligence is the process of bringing together all types of information a company has in a way that’s relevant to the question.

Background:

From stakeholders perspective, this business intelligence project is to really understand the amount of manpower that looking for barriers that is prohibiting us from achieve full inspection tasks compliance throughout the year.

In the past, if the team is unable to complete planned tasks. We will just do something else instead and never really follow the schedule if needed. We won’t track down what’s going on and what are the barriers that prevent us to complete all the tasks. For instance, when some auditors ask that why we didn’t complete that task which is scheduled for a particular week, we could not answer it and that resulted in a lot of “fat” in job planning part. The inspectors and technicians will be free if the job that is supposed to be done on that week has been put on hold (due to resources constraint, weather, urgent leave). We need record down what they have done if they carry out the job, and how many hours are lost due to break-in, resource constraint. (Lost manhours).

Furthermore, the old way the analyst does is just pull out the all-task compliance in one Excel dashboard without further break down. What I have done is instead making the task compliance a monolith, by using the interactive and granular functions of Tableau, I am able break down the status into unplanned tasks status, planned task status, field completion status, report completion status and etc which provides a more insights to management.

What I have done is I have created a Tableau dashboard which glues tasks compliance and barrier hours together so that people can understand the relationship between manpower, lost manhours, barriers, and our status compliance. This dashboard help managers to make better decisions by really providing in depth insights of what happening on the ground and the most significant reasons that prevents us to complete the planned tasks. And thru this insights, our management were able to take appropriate actions which results in an average reduction of 20% barrier hours on average for the next 3 months and save up to 50K SGD per month.

After the successful implementation of Tableau at my site, I am asked to drive this Tableau applications across all 4 sites in whole SG. To me that is most challenging part, because I have to teach my teammates to perform data cleaning in data source, use Tableau and set up path to my tableau template to feed in the data accordingly. The teammates where I have does not even have good excel skills let alone skill to perform data analysis. I encouraged, very patient, meticulous when doing the coaching and they were able to complete their responsibilities as mine. We were able to complete this project successfully on deadline and we are able to show case our mutual effort in one dashboard.

**Strength**

Even though OnG is an well established industry, innovation is still very important for us to march forward.

S: There is a scenario we need to do radiography for piping sulfidation corrosion (a special kind of corrosion). Radiography is like X-ray but instead of X-ray we use gamma ray as source. Normally this kind of equipment, we cover the piping with film, let it expose to radiation source to get the profile of internal condition of piping. For accessible locations, it is very easy to set up, however we need access like scaffold for piping which has higher altitude 5m 10m. Money and resources are wasted.

T: When I visited vendor workshop, I found one special equipment which inspires me. It is a telescopic pole with camera mounted on top. I came out with an idea that I can buy 2 telescopic pole and mount source on one pole, another with special holder which can hold the film. We can actually save up some dollars for scaffold.

A: I have designed and worked with my team with a welded metal frame to mount these 2 things. It took us several attempts to come out with a design where it is able to hold these equipmets.

R: As a result, our specially designed equipment has been used in field successfully. We are able to save 10k dollar from preventing scaffold erection for this particular task. In our daily operation we need a lot of radiography support and I foresee that we can achieve more cost saving in the near future.

**Weakness**

Well, every coin has two sides, one must always have their own weakness, I was never confident with public speaking-which as you know, can be a hindrance in the workplace. When I realized this was a problem, I took the initiative to become the presenter in several meeting in school and not afraid to voice up during business meeting. As a result, I was able to overcome my fear and since then, I have cohosted an engineering forum for more than 200 people and presented the ML to 50 people in face to face meeting and receive good feedback from the leadership team. Regarding this, I still find public speaking challenging but enjoyable for me now.

**Mistake**

There was a scenario when one of inspection supervisor who is managing different unit went on leave for 2 weeks. Our managed need to find volunteer to carry on his workload, I accepted his ongoing tasks as in I didn’t want to seem like I couldn’t handle extra work without checking out my schedules. When we have break-in tasks (because equipment leak and need manpower to monitor the equipment, I don’t have enough manpower to meet deadlines for planned tasks. I admitted this issue to my manager, we need to get extra manpower from different site. For instance, I will use application to note down my daily, weekly tasks to be done and indicate important project deadlines and where I am in the process at any given time and I am able deliver the results to my supervisor according to what I promised, I will also think carefully and thoroughly before giving any promise.