

PharmaSUG 2021 : Paper AI-049  
**How I became a Machine Learning Engineer from a Statistical  
Programmer**

Kevin Lee, Genpact

### **ABSTRACT**

One of the most popular buzz words nowadays in the technology world is “Machine Learning (ML).” Most economists and business experts foresee Machine Learning changing every aspect of our lives in the next 10 years through automating and optimizing processes. This is leading many organizations to seek experts who can implement Machine Learning into their businesses.

The paper will be written for statistical programmers who want to explore Machine Learning career, add Machine Learning skills to their experiences or enter a Machine Learning fields. The paper will discuss about personal journey to become to a Machine Learning Engineer from a statistical programmer. The paper will share my personal experience on what motivated me to start Machine Learning career, how I started it, and what I have learned and done to be a Machine Learning Engineer. In addition, the paper will also discuss the future of Machine Learning in Pharmaceutical Industry, especially in Biometric department.

### **WHAT IS MACHINE LEARNING?**

Machine Learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. The key of the Machine Learning definition is “**without being explicitly programmed**”. What it means is that the machines will learn new skills by themselves without programmers adding new rules for new skills. In other words, Machine Learning enables the computer to solve the problems with data and algorithms much like a human does.

### **WHY MACHINE LEARNING IS SO POPULAR NOWADAY**

The machine learnings or AI is predicted to revolutionize all the industries, especially healthcare industry. First, the Machine Learning can help us to solve a lot of complex business problems that we have not been able to solve before.

Secondly, Machine Learning can be also very cost effective since it can automate a lot of process. Andrew Ng, the founder of Coursera and previous chief data scientist of Baidu said, “**Pretty much anything that a normal person can do or think less than 1 second, we can now automate with AI**”. Machine Learning along with robotic is expected to automate a lot of human labors. According McKinsey, **as many as 375 million workers (14% of global workforce) may need to switch jobs due to automation by Machine Learning and AI.**

Due to its cost-effective potential and ability to solve complex problems, many companies and thought leaders are considering Machine Learning as the next industrial revolution. More and more businesses are implementing Machine Learning to innovate and lead the next industrial revolution. The companies like Google and Facebook target Machine Learning/AI as their priority, moving from mobile.

### **THE CURRENT MACHINE LEARNING IMPLEMENTATION**

Machine Learning is being used more than we realized. The followings are Machine Learning implementation in our daily lives.

- Voice Recognition System – Siri, Alexa, Google Home
- Recommendation – Amazon, Netflix, Spotify
- Customer Service – Online chatting (e.g., Chatbots)
- Cashless store – Amazon GO
- Autonomous Vehicles – Tesla, Google
- Image recognition - CT scans
- Face recognition

### **WHAT IS MACHINE LEARNING ENGINEER?**

Machine Learning (ML) Engineer is one who make machine learning models to implement in the production. Especially, ML Engineer can scale ML application with the large amount of data. Therefore, ML Engineer works with Data Scientist who can find the business solution with data and analytics and communicate the insights with business stake holders. ML Engineer also work with Data Engineers who can build the data pipelines, data storage and feature engineering.

In many occasion, ML Engineer should be able to conduct data preparation, and build the models and deploy the models in the production like an experienced Statistical programmer knows how to prepare SDTM, ADaM, Define.xml, TFL to the submission.

## **SKILLS THAT MACHINE LEARNING ENGINEER NEED TO LEARN**

There are many skillsets that ML Engineer or Professional job requires, and here is a sample job description of Machine Learning Engineer at Amazon.

- Graduate degree (MS or PhD) in computer science, engineering, mathematics, or related technical/scientific field
- 5+ years of professional experience in a business environment
- 3+ years of relevant experience in building large scale machine learning or deep learning models and/or systems
- 1+ year of experience specifically with deep learning (e.g., CNN, RNN, LSTM)
- Experience in using Python or other programming languages

Typical ML Engineer job description requires below skill sets.

- Strong Programing experience in Python, Java, or C++
- ML modeling experience in CVM, Logistic Regression, Regression, Decision Trees, Random Forest, K mean Clustering,
- Deep Learning experience in CNN, RNN, NLP.
- ML package experience in Scikit Learn, TensorFlow, Keras, PyTorch
- Cloud Computing environment in AWS, AZURE, Databricks, IBM Watson and Google Colab
- Database experience in Hadoop, Data Warehouse, Data Lake, NoSQL, Relational Database
- MLOPs experience in data pipelines, feature engineering, ML model selection/training/validation, and finally to the deployment in the production.
- Excellent communication and presentation skills

Not all ML Engineers have all those skill sets, and most of ML Engineers will learn the required skill sets on their jobs if they don't know. Most importantly, many skillsets for ML jobs will change because ML technologies continue to change and improve in a speed that not everyone can keep up with, and ML Engineers need to learn as fast as ML technology reinvents itself.

A great news for those who wants to learn Machine Learning is that there are a lot of ways to start.

- PhD or Master's degree programs in Universities / Colleges. One of the hottest major in college nowadays is Computer Science or Data Science. And many of degree-courses are also offered on-line.
- ML Certificate programs in college. The most popular certificate program is Machine Learning certificate program. A lot of major universities currently provide 3 to 9 months ML certificate program.
- On-line courses. Many on-line ML courses are available in Coursera, EdX, Udemy, Udacity, Data Camp and more. Those courses do not provide degree, but they are as good as any college courses.
- YouTube. They are many great YouTube Machine Learning courses and video in <https://youtube.com>.
- GitHub. Many ML models and implementations are posted in GitHub in <https://github.com/>. GitHub also provides self-studying materials and codes. So, many beginners could download sample codes and data, and they could practice and run ML models in their own environments.
- Kaggle. Kaggle (<https://www.kaggle.com/>) is ML practices and competition environment. Kaggle provides a great way to test your ML knowledges and build ML experiences. If you become Kaggle Grandmaster or master, you will be recognized in ML community.
- ML Books. They are many great ML books that you could learn.

Many of ML professional do not have computer degree nor PhD degree. They usually learn from videos, books, on-line courses, and each other.

## **HOW I LEARNED MACHINE LEARNING**

The first time when I was exposed to Machine Learning is when I was working with Big Data Company, and I realized that Machine Learning is the next big technology, which will lead the next innovation.

At first, I started as a curiosity. I did not mean to change my career, but just wanted to know about ML technology. And since I was working from home, I could save about 2 hours per a day for the commute, so I spent that much time in learning Machine Learning.

I started in two directions : **Machine Learning Algorithm and Python Programming.**

### **Machine Learning Algorithm**

For Machine Learning Algorithms, I started on-line Machine Learning courses in Coursera. Below are the courses that I took.

- Machine Learning at Stanford University

- Neural Network & Deep Learning at DeepLearning.ai
- Improving Deep Neural Network : Hyperparameter tuning, Regularization & Optimization at DeepLearning.ai
- Convolutional Neural Network at DeepLearning.ai
- Sequence Models at DeepLearning.ai
- Structuring Machine Learning Projects at DeepLearning.ai
- AI for Everyone at DeepLearning.ai
- AI for Medical Diagnosis at DeepLearning.ai
- AI for Medical prognosis at DeepLearning.ai

I could have attended college level ML courses at college and graduate schools, but on-line ML courses are as good as college degree courses. Especially, I learned a lot from Coursera ML courses taught by Andrew Ng, Sandford University professor. I have learned pretty much same knowledges and concepts from these ML on-line courses, and they were a lot cheaper than colleges. The downside is that I did not get personal supports like in college from professors.

### **Python Programming**

While I was taking on-line ML courses, I also studied Python Programming. Python Programming is the best computing language for ML models and implementation. I am hands-on person, and I wanted to use Python programming to train, develop, validate, and deploy ML models. Since I loved coding, I started Python programming right away with books and its examples. Below are the books that I studied with.

- Python Crash Courses
- Python for Data Analysis
- Feature Engineering for Machine Learning
- Hands-on ML with Sci-Kit Learn & TensorFlow
- Python Machine Learning
- Apache Spark Deep Learning Cookbook

I should have taken Python programming courses since they would probably give me better starting points such as Python programming platform (e.g., Jupyter notebook) and differences between SAS and Python. The books have their own GitHub accounts, so I was able to download the Python codes, run and play with Python codes.

After understanding ML models and programming, I started to use them in my works and share with others. And nowadays, I am teaching Python programming and ML models at Corporates, conferences and University.

### **TRANSITION FROM STATISTICAL PROGRAMMER TO MACHINE LEARNING FIELDS**

A transition from Statistical programming to Machine Learning is not as difficult as it looks. However, a transition could be a couple of steps. From my personal experience, I transitioned into Data Science role first from Statistical Programmer. There are a lot of commons between Statistical programming and Data Science – Programing, Statistics and Data. One major difference could come from programming language. Data Scientists could use many open-source programs like SQL, R, Python, Java while Statistical programmers mainly use SAS®.

Once I became a Data Scientist, I learned ML models and Cloud Computing, then ML field opportunities got eventually open for me.

If you obtain the advanced degrees in ML at graduate school, then you could jump into ML fields much easier.

### **WHAT IS A SALARY FOR MACHINE LEARNING ENGINEER?**

Based on Indeed Salary guide, the average salary of Machine Learning Engineer is \$150K while as that of SAS programmer is \$115K. The highest paying companies are offering more than \$200K. And its demand has been increased almost 10 times since 2012.

### **MACHINE LEARNING IMPLEMENTATION IN PHARMACEUTICAL INDUSTRIES**

Machine Learning is believed to have a huge impact in pharmaceutical industry since 1/3 of data comes from healthcare industry. Its implementation already started in many pharmaceutical companies, and more and more pharmaceutical companies utilizes ML technologies. Many pharmaceutical companies try to utilize AI/ML to automate and speed up the following areas.

- Drug discovery
- Drug candidate selection
- Supply Chain optimization
- Medical image recognition
- Medical diagnosis
- Optimum site selection or recruitment
- Data anomaly detection
- Personalized medicine

- Medical coding
- Sales and Marketing Optimization
- Pharmacovigilance
- Drug Development

### **SO, WAS IT WORTH TO BE A MACHINE LEARNING ENGINEER?**

Based on the prediction by Forbes, the global machine learning market was valued at \$1.58B in 2017 and is expected to reach \$20.83B in 2024, growing more than 40% annually. The demand of ML professionals will get much higher as the ML market grows.

Even if Machine Learning market will grow exponentially, Statistical Programming market also looks very strong in near future. I don't believe that Statistical Programmers should transition out of Statistical Programming because of a short of Job Market. If Statistical Programmers want to expand their roles and to lead with data and analytics, ML is worthwhile to add in our skillsets. Could we imagine we could do both Statistical Programming and Machine Learning?

### **CONCLUSION**

Machine Learning is believed to have a huge impact in Pharmaceutical Industries and there will be a lot more demand and opportunities in ML fields. Having additional ML skillsets will be great opportunities for statistical programmers. Whether we transition to Machine Learning Engineers or stay as statistical programmers, if we are open to ML implementation in Biometric Department and work with ML technology, we can lead the next innovation of pharmaceutical industries with data and analytics.

### **REFERENCES**

- GitHub in <https://github.com/>
- Kaggle Competition in <https://www.kaggle.com/>
- TensorFlow, <https://www.tensorflow.org/>
- Sci-kit Learn, <http://scikit-learn.org/stable/>
- Coursera in <https://www.coursera.org/>
- Machine Learning Engineer Salary Guide in Springboard in <https://www.springboard.com/blog/machine-learning-engineer-salary-guide>
- Indeed Salary Guide for Machine Learning Engineer and Statistical programmer in <https://www.indeed.com/career/sas-programmer/salaries> and <https://www.indeed.com/career/machine-learning-engineer/salaries>
- Forbes (Roundup of Machine Learning Forecasts and Market Estimates, 2020) in <https://www.forbes.com/sites/louiscolombus/2020/01/19/roundup-of-machine-learning-forecasts-and-market-estimates-2020/?sh=39ae5ff25c02>

### **CONTACT INFORMATION**

Your comments and questions are valued and welcomed. Please contact the author at

Kevin Lee  
 AVP of AI and Machine Learning Consultant  
 Genpact  
 Kevin.lee@genpact.com

### **TRADEMARKS**

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries.

® indicates USA registration. Other brand and product names are registered trademarks or trademarks of their respective companies.