MERIT Internship report (domestic company) Felix Julian Brandenburg (37-167086)

I had the honor of doing my internship at NextGem Inc., a japanese biotech venture company trying to implement futuristic data science technologies (deep learning, machine learning, etc.) in order to create high-impact solutions. In this report, I would like to first briefly introduce the reader to the industrial background as well as give a brief introduction of NextGem Inc. Afterwards, I will speak in more detail about my every day work during my time at NextGem Inc from November 2018 to March 2019.

The field of biotechnology is quite interdisciplinary, where traditional biology (enzymes, cells, DNA, pathologies, etc.) meets novel cutting-edge technologies from other fields (e.g. engineering). As such, biotechnology companies, using their vast biological know-how and laboratory capabilities, often focus on creating new applications.

I feel that this is in essence everything that MERIT stands for: striving to become an industrial or technological power-house by collaborating between interdisciplinary fields (i.e. physics- and chemistry-sided) and thus achieving unprecedented success.

NextGem Inc. is a recently started biotech venture company with its roots being in Kobe. The company executives are all graduates from Kyoto University from various fields, ranging from electrical engineering to medicine and biology. As such, the NextGem team is well-versed in biological, medical, as well as engineering vocabulary, allowing the team to quickly tackle new problems from various angles in a professional manner. Due to their close collaboration with various companies and hospitals in the Kobe area, NextGem is in the position to create a profound and unique data-base, an optimal environment for aspiring data scientists who wish to have a positive impact on the quality of life of a wide range of human beings living in Japan (and soon also outside). NextGem also does biological and medical experiments in their wet-side facilities based in Kobe, meaning that unlike other biotech companies, NextGem has state-of-the-art tools to directly verify their data science based predictions in their laboratories. As part of their dry-side (i.e. data science) team based in Tokyo, I had the opportunities to visit their facilities in Kobe. Their Kobe based laboratory is quite international and interdisciplinary. More specifically, NextGem works in close collaboration with various other laboratories throughout the world, with one of the more prominent ones being Stanford University School Of Medicine. Its close ties with Stanford University School Of Medicine allowed NextGem to obtain various patents in the field of Stem Cell technologies, giving them the edge over rival biotech companies. As such, NextGem's vision is to use their pristine understanding of medicine and biology, e.g. newest understanding of how human stem cells work, in combination with machine learning and deep learning techniques, in order to predict a patient's likeliness to be befallen by specific illnesses or conditions.
The goal is hereby to create an application which can be a useful tool for nowadays doctors, since machines work and think differently than humans. A doctor can use his experience, knowledge and common sense in order to diagnose a patient with a specific condition, which is generally sufficient if the doctor is competent enough. On the other hand, machines look and study the patients parameters, which can often be in the hundreds, in order to find correlations and thus predictions regarding future dangers. More concretely, we train the

machine with a training data set in which it studies which parameter patterns may cause what kind of conditions. After the training step (which may involve thousands of iterations), the machine should be in a position to make its own prediction for future, new parameters. One attractive point of using machine learning and deep learning in the medical field is that unlike doctors, machines work always 100% precisely and are often able to detect dangers which doctors are unable to sense. An example hereby is detecting potential cancer lumps by looking at the patients scanned image data, or studying a patient's heart rate pulse images in order to find potential dangerous conditions. While hereby a doctor would arguably need years of experience in order to be able to detect subtle signs of illnesses, the machine will become more precise in little to no time.

During my time as an intern at NextGem Inc., I was responsible for programming a prediction model which could help doctors in evaluating the health of new-born babies. For creating the machine learning model, I purely used python as the programming language. Python offers a wide-range of freely available data science packages, from which I mainly focused on the scikit-learn and tensorflow (now as keras) packages. More precisely, the data mainly comes in two types: numerical data (linear or binary) or image data (jpg or csv). A patients numerical data can represent body parameters (age, height, weight, etc.) while binary data can be used as flags for specific types of conditions or medicine consumption (e.g. taking medicine A?: 0 => no, 1 => yes). For analyzing purely numerical data sets, it is often enough to simply rely on machine learning algorithms. Hence I spent a lot of time on finding the ideal machine learning model (and algorithm), in terms of precision and running time (relevant once running high volumes of data). In order to find the ideal algorithm, one had to carefully study the data set first. Data sets with high amounts of binary data turned out to be bestanalyzed with decision tree algorithms (e.g. ExtraTree Forest or Random Forest), while datasets with lots of linear, constant data was more suitable for linear regression models. By choosing the right algorithm, I was able to increase the prediction model of the NextGem software for new-born babies to the high 90%.
In the second half of my intern at NextGem, I was more focused on creating a neural network with which I would be able to analyze imagedata. The package I hereby chose was Keras, which is basically a more simpler user interface of the highly complex and functional Tensorflow package. Tensorflow lets us create CNN (convolutional neural networks), which are able to learn from input image data,re-calibrate its networks weight and bias paths, and in the end make predictions for new, unrelated images. Based on this principle, I was able to let the CNN study a pregnant mothers heart rate flow chart, in an attempt to detect and inform the doctor preliminarily incase of abnormalities. Unlike conventional machine learning programs, deep learning (CNN) programs (mostly image recognition applications) require more computational power,and in the case of high data volumes, should ideally be performed on a supercomputer or acloud computer (e.g. services by Google, Microsoft or IBM). As such, I also had the chance to learn about general big data concepts, i.e. how to create and maintain an infrastructure which lets us operate effectively for even high data volumes.
Overall, my time at NextGem was a truly unique experience, and while working as a part of their highly interdisciplinary and innovative team, I was able to gain new insights about the industry as well as academics. NextGem Inc. is the ideal brain-storm hub for University of Tokyo graduate students, as they like people with unique backgrounds and also give you the

opportunity to express your viewpoint and technical knowledge. I learned how to formulate my technical knowledge (physics and electrical engineering) in a way such that even people from other backgrounds were able to understand me, and vice-versa, I had the opportunity to learn from Medical Doctors as well as University of Tokyo Economics PhD students, and was able to adapt to their style of thinking. Incase you are interested in working in a highly flexible,motivated, innovative and competent environment, I definitely recommend joining a startup or venture company.

I would like to deeply thank my supervisors, Mr. Masakazu Nakajima and Mr. Isao Miyatsuka, from NextGem Inc. for giving me the opportunity to have this internship at the first place. They were kind and patient enough to introduce me to their companies technologies, while at the same time giving me the time and space to develop and implement my own ideas.
I would also like to thank Dr. Masanori Miyanishi from NextGem Inc. for introducing me to the general concepts of biotechnology.
Also, I would like to express my sincere gratitude for my supervisors at the University Of Tokyo, Dr. Takashi Yatsui, Dr. Takuzo Aida and Dr. Masashi Kawasaki, for their continued advice and support.
Lastly, I would like to thank the MERIT program itself, for giving students this wonderful opportunity to gather first-hand experience within academics and the industry.