Questions for the Industrial Master's Programme in Bolt

Name: Xinjian Zhang
Program: MSc Software Engineering
Email: xinjian.zhang@ut.ee

Bolt

Preferred topic: 2. Machine learning models to monitor metrics and data quality.

---

1. What drew you to the specific topic you are applying for?

- **Relevance to Previous Experience**: Since I studied for my bachelor's and now my master's, I have been involved in some courses and projects related to the machine learning model. In my previous project, I noticed the importance of data quality and its effect on the model's performance. So, I am interested in this topic, as my next step in my ML study.
- **Passion in Research**: In the age of AI, I've always been fascinated by how ML can solve practical problems and work with other fields. Meanwhile, the data and statistics are critical as well. This topic offers a chance to dive deeper into the research in these areas.
- **Importance in Industry**: Data quality and metrics are crucial for businesses' decision-making. By working on this topic, I can learn how to monitor and improve data quality, a valuable industry skill contributing to the company's success.

2. How do you see your work in this program contribute to your broader career goals, and how do you plan to leverage this experience to grow and develop professionally?

This program presents an ideal chance to acquire hands-on industry experience, which is crucial for advancing my career and will significantly support my academic endeavors. It does so in several ways:

- **Hands-On Experience**: Working with a company like Bolt offers me a chance to tackle real-life challenges and gain hands-on experience. This will help me apply what I've learned in a practical way, which is key for my future job. By dealing with actual projects, I can better understand how to use my knowledge to solve problems and develop skills that will be valuable in my career.
- **Networking Opportunities**: Working with professionals in the field is a chance to make valuable contacts and learn from them. I see networking as crucial for my career growth, as it allows me to link up with experts and others in the industry.
- **Academic Mentorship**: Working with a supervisor will provide valuable insights and guidance, improving my academic skills and knowledge.
- **Skill Enhancement**: This programme will help me enhance my machine learning, data science, and problem-solving skills. Which is essential for data-related careers.

3. Describe a machine learning project you've worked on that you are proud of. What was it about? What were the challenges you faced, and how did you overcome them?

One of the machine learning projects that I'm particularly proud of involved developing a 3D convolutional neural network (CNN) model to analyze medical imaging data. This project aimed to improve the accuracy of predicting certain medical conditions from 3D medical images.

- **Objective**: The goal was to use 3D CNNs to process and analyze medical imaging data, aiming to identify and predict specific conditions with higher accuracy than traditional methods. This involved working with .npz files containing volumetric data captured by medical imaging techniques.
- **Approach**: The project utilized several key technologies and techniques:
- **Data Preprocessing**: Loaded and processed imaging data from .npz files. This involved selecting specific segments of the images to focus our analysis on regions of interest, enhancing the model's ability to learn relevant features.
- **Model Architecture**: Designed a Sequential model in Keras, incorporating Conv3D layers for volumetric data processing, MaxPooling3D layers for feature downsampling, and Dropout layers to prevent overfitting. The model concluded with Dense layers for classification.
- **Normalization and Reshaping**: The imaging data was normalized to a 0-1 scale and reshaped to fit the model's input requirements, focusing on a sub-volume of the original data to improve processing efficiency and model accuracy.
- **Challenges**:
    - **Data Volume and Quality**: One major challenge was handling the large volume of 3D data, which required significant preprocessing to ensure quality and relevance for the model.
    - **Model Complexity and Training**: Designing a model complex enough to capture the nuances of 3D medical data without overfitting or becoming computationally infeasible to train.
- **Solutions**:
    - **Efficient Data Handling**: By focusing on sub-volumes of the imaging data, we could manage the data volume more efficiently, allowing the model to train on the most relevant features.
    - **Model Optimization**: Experimented with different layer configurations and hyperparameters to find the optimal balance between model complexity and training feasibility. Additionally, the use of dropout layers helped mitigate overfitting.

In summary, through this project, I gained invaluable experience in 3D data processing, model optimization, and the practical application of machine learning in healthcare.

Last words

In previous and current academic experience, I utilize Python, SQL, and Jupyter Notebooks in many courses, projects, and Python-related libraries. For example, libraries such as Numpy, Pandas, and Scikit-learn are frequently used in are covered in this semester's business data analysis course.

The first time I got to know the company Bolt was from the news many years ago when it was still called Taxify. I'm amazed at its transportation solutions and rapid development. In my last study, I participated in the Digital Product Management Industrial Program, working with Bolt to develop a prototype based on business travelers. In the process, I learned about the diversity and innovation of the Bolt team. Therefore, I especially hope to have the opportunity to cooperate with Bolt to continue my learning and career development.