

# MyungWon Lee

• Edinburgh, United Kingdom • [johnnymwlee@outlook.com](mailto:johnnymwlee@outlook.com) • [in/johnnymwlee/](https://in/johnnymwlee/)

## Summary

The most bright and enthusiastic person you will ever meet. Passionate with analytical mindset on various sets of data and a machine learning researcher who preserves innate curiosity to grow and add value as a strong researcher.

## Education

<b>University of Edinburgh</b>   <i>Doctor of Philosophy in Statistics</i> • Supervisor: Dr. Daniel Paulin • Research Interest in Bayesian statistics, meta learning, machine learning and computational statistics	Sep 2022 – Incoming
<b>University of Edinburgh</b>   <i>Bachelor of Science (Honours) in Mathematics &amp; Statistics</i> • 1 <sup>st</sup> Honours Degree • Thesis on “Semismooth Newton Augmented Lagrangian Method for Solving Lasso Problems with Implementation in R”	Sep 2016 – May 2022
<b>Seoul National University</b>   <i>Bachelor of Science in Statistics</i> (3 <sup>rd</sup> Year Exchange Student)	Aug 2018 – Jun 2019

## Experiences

**LG Chemical Patent Analysis & Samsung Securities News Curation Industry Research** Jun 2021 – Aug 2021  
*Data Mining Laboratory Intern, Seoul National University, Seoul*

- Analysed 1,198 solid-state battery cell patent documents to classify sentences into problems or solutions by building unsupervised learning models such as **BERT**, **char-CNN**, **RNN** with **NLTK**, **Scikit-Learn** and **TensorFlow**
- **Increased 20% model accuracy (F1-score)** by performing text pre-processing/post-processing using **RE** package to cleanse data
- **Decreased parsing error rate from 35% to 3%** by suggesting alternative parsing method for SEC Edgar 10-K reports using **BeautifulSoup**, **RE**, **Gensim** in Samsung Securities News Curation Project team

**Financial Value Insight Data: ETF & Glassdoor Research** Jun 2019 – Aug 2019  
*Data Mining Laboratory Intern, Seoul National University, Seoul*

- Integrated large-scale web-crawling on *morningstar.com*, *finance.naver.com*, *comp.fnguide.com* | *glassdoor.com* & *jobplanet.com* using **Python** packages such as **Selenium Webdriver**, **BeautifulSoup**, **RE** and **Pandas**.
- Identified correlation between company's annual profit and anonymous company's reviews through natural language processing
- Nominated by team leader as a presenter in the weekly lab seminar on behalf of the Glassdoor research team.
- Gathered and recalculated the missing financial fundamentals for 435 Exchange Traded Fund from Korea Exchange. Involved in building database schema to store collected data using **SQL & Django** framework.

**DataAssembly** Feb 2018 – Aug 2019  
*Co-founder & Consultant, Buchanan Institute, University of Edinburgh, Edinburgh*

- A student-led data consultancy **DataAssembly**, providing pro-bono data analytics to small business & third sector industries through students' engagement in the real-world problem using **R** and **Python**.
- Gained invitation as a representative of **DataAssembly** to DataFest Event 2018 organised by the Scottish Government.
- Acknowledged, fulfilled, and identified the needs of team members and clients, approaching clients from door to door and guiding students from non-statistical backgrounds with structured curriculum and fair opportunities.

## Projects

**R package development for “Semismooth Newton Augmented Lagrangian Method for Solving Lasso Problems”**— *Final Year Project, University of Edinburgh, Edinburgh* Sep 2021 – Jun 2022

- Developed a fast and highly efficient algorithm to solve Lasso regression with implementation the algorithm in **R** with **C++** and **showed lower mean squared error** compared to **glmnet**.
- Improving the algorithm on its smooth starting and planning to publish on CRAN. ([GitHub link](#))

**Newsletter Mall Project** — *Self Designed Project, Seoul* Mar 2021 – Aug 2021

- Identified the need of rapid supply of accurate news in the Korean stock market to solve unexpected risk and lack of information
- With Kiwoom Securities trading **REST API**, Naver **OpenAPI**, and telegram **API**, developed a real-time bot sending out newsletter. Developed an auto trading algorithm based on the information crawled.
- Mecab tokeniser from KoNLPy package to pre-process and censor unwanted AI generated news and advertisements.

**Pattern Analysis of New York Airplanes Data Research** — *Seoul National University, Seoul* Sep 2018 – Dec 2018

- Analysed 120 million records of flight arrival and departure data of commercial flights within the USA using **R** and **SQL**
- Provided graphical and visual summary of the best **time of day** / **day of week** / **time of year** to fly to optimise delay hours with supervised learning such as linear regression, logistic regression, ridge regression, lasso regression, linear discriminant analysis, quadratic discriminant analysis, decision trees, random forest, boosting and neural networks

## Skills

Machine Learning: Python/R | Front End: HTML/CSS/JavaScript | Pitching | Presentations | Proactive | Teamwork | TechSavvy  
Microsoft Office Specialist–Masters (**Jun 2017**) — Word Expert | Excel Expert | PowerPoint Core | Access Core

**Military Service** — *Republic of Korea Army* Aug 2019 – Mar 2021  
Honourably discharged from 36th Infantry Division, Department of Intelligence.