DSCI 542 Lab 2

Creating a Blog Post for Your Audience

Table of contents

Lab Assignment Instructions

In this course, we are using a labor-based grading approach. This means that each task in the lab is graded based on completion rather than evaluation of quality. If a question is completed according to the specifications provided such as meeting the word limit, following the required format, or staying within the allotted time you will receive full credit for that question.

When submitting your lab, please be sure to:

- Include a link to your GitHub repository containing your work as well as a link to your published blog post.

  https://github.ubc.ca/mds-2025-26/DSCI_542_lab2_jfc25

https://github.com/Jacob-F-Cann/542_Lab2_Blog_Post

https://jacob-f-cann.github.io/542_Lab2_Blog_Post/

- Submit a the `.qmd` file along with a rendered version (either `.html` or `.pdf`) of your lab to Gradescope.
- On Canvas, submit a link to your blog post for peer review.

Question 1: Understand your audience

rubric={completion:5}

In this lab, you will be tasked to writing a data science blog based on the topic you chose in lab 1.

As mentioned in Lecture 2, an audience persona is a brief written description of a specific person in your target audience, who might be real or imagined. It should include the following information about the person:

- Their **experience**. How much experience do they have with your subject, and of what kind?
- Their **motivation**. Why do they want to learn about this subject? If they have any specific goals, what are they?

For example, here are three personas for potential audiences of an explanation of quantum computing:

- Mike is a third-grade student. He has fairly little experience with any computers at all, let alone quantum computers. But he suspects that computers might be able to help his with tough homework problems, and wants to know if this is true of quantum computers.
- Nadine is a fifteen-year-old high school student. She is familiar with computers; most commonly she uses her iPhone to socialize, study, and take photoes. She has never heard of quantum computers before, but likes to impress her friends with cool facts about things she learns.
- Alex is an undergraduate computer science student interested in machine learning and artificial intelligence. They taken courses in chemistry and physics, and knows the basics of how quantum computing differs from classical computing. They are curious about the kinds of problems that quantum computing could be used to solve, and how soon the technology might become commercially available.

Write an audience persona for the target reader of your blog that is **3-5 sentences long**. Additionally, propose a **minimum of 3 different titles** for your blog. During the lab, we encourage you to exchange the audience persona, the titles, and the abstract with your peers and get their initial feedback.

Persona:

Bob is a PhD Student at UBC beginning his investigative research for his thesis on predictive models for battery capacity lifetime degradation. He intends on taking a data science approach to the problem instead of classical physics-based modelling simulation methods, and is looking for blogs, articles and texts covering the subject to generate ideas for his thesis proposal. Bob has a Master's in Chemical Engineering specializing in battery chemistry as well as a Bachelors degree in chemical engineering. He has taken a lot of online introductory and advanced courses for machine learning basics and applications in preparation for his targeted PhD research. Data science and machine learning is definitely more unfamiliar than pure chemistry, physics and battery design for Bob, so he has a lot to learn still despite his online courses.

Title 1: Machine Learning for Battery Lifetime Prediction

- Informative but too bland for a title to capture a reader's attention

Title 2: When will my Electric Car Battery Die?

- Good hook but too mysterious to be informative

Title 3: The Cure to Range Anxiety - Battery Lifetime Prediction with Machine Learning

- Title 3 is the strongest in my opinion. It has a hook and is informative of the content

Question 2: Write a data science blog

rubric={completion: 30}

Guideline to get started

I highly recommend you check out this guideline from TowardsDataScience.com, a well-known publisher for data science blogs.

Types of blog post in data science

There are many different types of blog posts that can be written in the field of data science. Some common types of blog posts include:

- **Tutorials**: These posts provide step-by-step instructions or guidance on how to perform a specific task in data science, such as using a machine learning algorithm or creating a data visualization using a specific package/language.

- **Explanations**: These posts aim to explain complex concepts or techniques in data science in a way that is easy for a general audience to understand. They might include examples or case studies to illustrate the concept or technique.
- **News and analysis**: These posts discuss recent news or research in the field of data science, and provide analysis or commentary on the implications or significance of the news or research.
- **Personal experiences**: These posts share the experiences of a data scientist or student learning data science, and might include stories about challenges and successes, as well as tips and advice for others interested in pursuing a career in data science.
- **Comparisons and contrasts**: These posts compare and contrast different tools, approaches, or concepts in data science, and might include pros and cons or recommendations for when to use each option.

Overall, there are many different types of blog posts that can be written in data science, and the type of post will depend on the goals and audience of the blog.

For your inspiration, this is a blog post that Nico Van den Hooff, a MDS alumni, wrote last year
Exploring the Most Popular Machine Learning and Deep Learning GitHub Repositories

Length

For this assignment, we suggest aiming for a word count of around 1200 (+ or - 300) words, not including code, captions, references, etc. Please consider the amount of time it will take your audience to read your post and keep this in mind as you plan and write your content. We recommend creating a brief outline for the blog first before diving in.

How to write your blog

In this question, you are asked to write your blog using Quarto.

Step 1: Login to Github

Step 2: Create your Quarto Blog.

Step 3: Start writing and finish! Push/publish your changes, and .

Step 4: Submit the blog's outline as well as **a link to your blog**. Your blog can be published either via Quarto Pub or GitHub Pages.

Your blog will be evaluated for completion based on the word count as well as the link to your blog functioning as expected.

GitHub: @Jacob-F-Cann

DSCI_542_Lab2 Repo

GitHub Pages for Blog

Some prompts to get you started (Optional)

- What is the main topic or concept you want to write about in your blog post? What makes it interesting or important to your audience?
- What type of data science blog do you want to write? (see different types above)
- What are the key points or ideas you want to include in your blog post? How can you organize these points to create a clear and coherent structure for your post?
- How can you use examples or case studies to illustrate the concepts or techniques you are discussing in your blog post? What are some real-world applications of the topic you are writing about?
- How can you make your blog post engaging and interactive for your audience? What types of media or interactive elements can you include to make your post more interesting and interactive?
- How can you conclude your blog post in a way that leaves a lasting impression on your readers? What are the main takeaways you want your readers to remember after reading your post?

Blog Post:

Why Should we Care about Battery Lifetime Prediction?

Batteries have become widely used in many industries including medical devices, automotive vehicles and energy storage systems. As part of a push for renewable energy solutions to combat climate change and the demand for better wearable and or implantable medical technologies, battery technology has accelerated in growth in recent years. Electric cars are more prevalent than ever, energy storage for renewable solar and wind energy is growing faster than ever and portable battery powered devices are sold by countless companies. The adoption of electric vehicles is divisive to say the least, and a major deterrent for people is "range anxiety". Range anxiety refers to when people alter or would alter their driving habits or trips due to fear of running out of charge in their electric vehicle battery. This can happen due to poor trip planning, or areas in countries with few electric charger stations where the potential for getting stranded is real.

For most day-to-day commuters with electric vehicles this is not a worry. However, consider additionally the cost of replacing an electric vehicle battery early. The cost could be a large proportion of the total cost of the vehicle itself, and moreover, you may have convinced yourself to buy an electric vehicle due to the savings on gas. Perhaps you finally reached your payback period on your electric car purchase providing savings on gas to offset the delta between the gas vehicle price and the electric vehicle price. All of those savings can go down the drain if your battery needs to be replaced all of the sudden after you hit your payback period…

What is the solution? It is easy - let's increase the lifetime cycle of electric vehicle batteries! Oh wait... this is not a trivial task... If that is not enough motivation for why this matters, maybe I should also mention that the probability of a catastrophic failure in the battery pack potentially leading to a fire or explosion, skyrockets when a battery degrades to a state of health (SOH) where it only has 80% of its original capacity to hold charge.

Batteries are complex physical systems, which demand analysis of data and design considerations in multiple displines of engineering inlcuding mechanical, materials, electrical and chemical. It turns out, that you can model a large proportion of the the complex behaviour of the internal battery components and physics, if you can accurately estimate the battery's current and future state of charge (SOC). The SOC is a function of time and current going in or out of the battery and is measured in amp-hours (Ah). It tells us how much charge (energy per-Volt) is available in the battery to do useful work. The main issue with modelling SOC over time is the fact that use of the battery affects the mechanics of SOC. A major issue with battery technologies is the decrease in SOC overtime due to use. This phenomenon is referred to as battery ageing, and the amount of complexity and interdependent mechanisms at play make it very difficult to estimate with standard physics equation paradigms. Predicting in a quantitative way when this degradation will occur and by how much, turns out to be an extremely difficult problem with many variables influencing the so-called battery capacity lifetime (or commonly just referred to as battery lifetime).

Battery Liftime Prediciton in Era of AI

One of the reasons machine learning and AI are such powerful tools in modern data analysis is their wide range of applicability to real-world problems across many different fields. One such field with extensive ongoing research for machine learning applications is battery design and technology. The analysis of the data centered around these technologies has therefore become increasingly important in recent years as well. Machine learning and AI techniques for predicting of battery system behavior with laboratory or field data is at the frontier of development of these technologies. It turns out that with enough voltage, temperature and current data of a battery charging and discharging (simulated battery use in a laboratory or collected in the field during use), these predictions become tractable with physics modelling frameworks and machine learning techniques.

Here is an example of real-world drive-cycle data for an electric vehicle, which could be used in a training datset for a machine learning model describing power consumption and regeneration. Note that voltage, temperature and current data would also be logged during this test. We can see that drive-cycle consumption patterns, and therefore SOC variation changes with the category of road type as well.

Figure 1 shows us the power consumption/regeneration in kW. When the vehicle is braking regeneratively the battery is charging and the power consumption is negative, whereas when

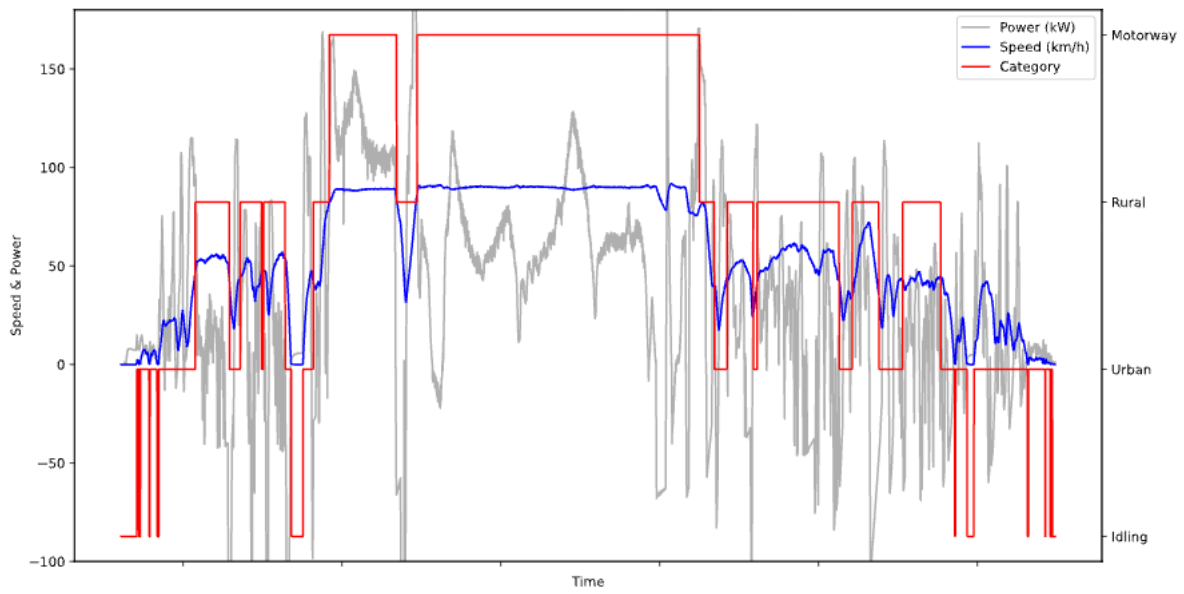


Figure 1: Deep Dive Analysis 1: Drive Cycle Assessment [5]

the vehicle is holding a constant speed or accelerating power consumption is positive and the battery is discharging.

The state of the art of battery lifetime prediction models, almost always involves machine learning today [1][2][3][4]. The models applied in this domain are time series in nature, and therefore the examples in the training and test data are causally linked through time dependence unlike models of other regression or classification applications. This comes with its own assumptions and challenges, however the battery lifetime prediction has become increasingly accurate and reliable consequently.

Here are some examples of what battery lifetime prediction with machine learning can tell us...

Figure 2 shows us how the SOC percentage decrease predictions change with charge-discharge cycling (use) for different ageing factors, relative to beginning of life (BOL) when the battery was manufactured.

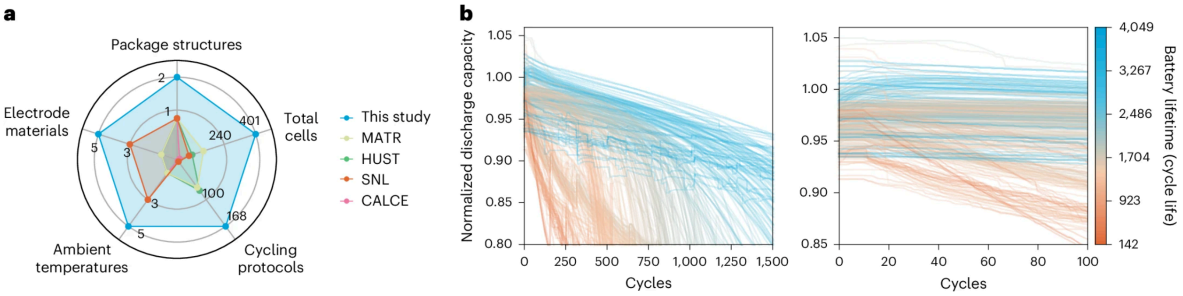
A similar cycle-life plot is shown below, but for predictions for different chemistry types:

We can see from Figure 3 that the physics-informed feature model predicts the cycle life for different chemistries with a reasonable spread of cycles, but some chemistries are harder to predict than others.

Another approach is to use formation data to build the model for predictions. Formation is the cycling done on the battery cell during manufacturing in order to stabilize the chemistry.

Fig. 1: Diverse ageing conditions covered in this study and complex degradation behaviours associated.

From: Battery lifetime prediction across diverse ageing conditions with inter-cell deep learning



a, We compare the broad coverage of different ageing factors in this study against typical datasets, including MATR[9], HUST[12], SNL[19] and CALCE[13,14,15]. **b**, The varying capacity degradation behaviours over long- and short-term cycles.

Figure 2: Battery lifetime prediction across diverse ageing conditions with inter-cell deep learning [1]

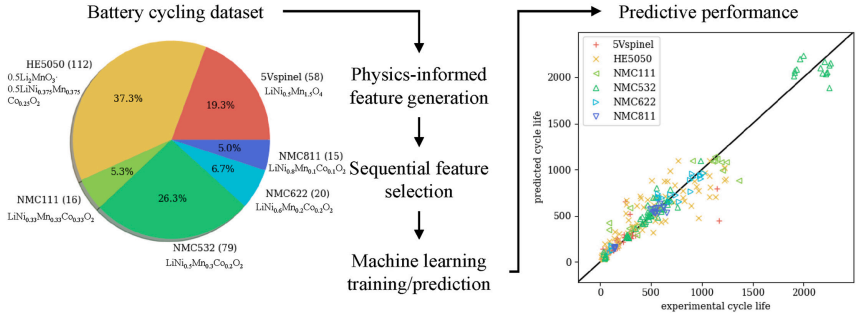


Figure 3: Feature engineering for machine learning enabled early prediction of battery lifetime [2]

Typically formation cycles are not counted as cycles before BOL, since the cycles are not actual end-user field use. Additionally, SOC calculations during formation can vary widely before stabilization occurs.

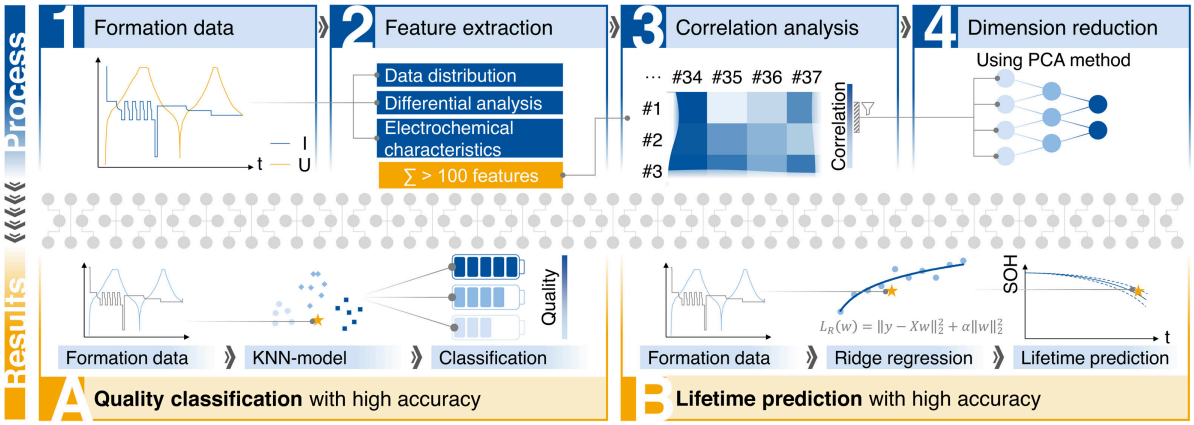


Figure 4: Machine learning for battery quality classification and lifetime prediction using formation data [3]

Figure 5 below also illustrates how deep learning can be used in the battery lifetime prediction machine learning pipeline with real-world driving data:

Problems with the Machine Learning Approach

The major problem in the field currently is the scarcity of high-quality data for training the models [1]. This is due to proprietary company data not being widely available and the cost of collecting the data as it requires intensive labor, capital and physical space to execute testing, in comparison with data collection for other applications. If the scarcity of data problem can be solved in the future, we will without a doubt see advances in machine learning for battery lifetime prediction. Another problem has to do with the fact that obtaining this real-word, large, high-quality training datasets has time and capital limitations on it. Evolving stable battery chemistry formulations, that are safer with a greater lifetime are constantly being researched and developed. The adoption of new chemistries will require the collection of new testing data and field use data in order for Machine Learning models to catch up with the new technology. In this sense, successful machine learning models for lifetime prediction will likely always lag behind battery chemistry innovation and this may slow the progress of the research and technological development.

References

1. Battery lifetime prediction across diverse ageing conditions with inter-cell deep learning

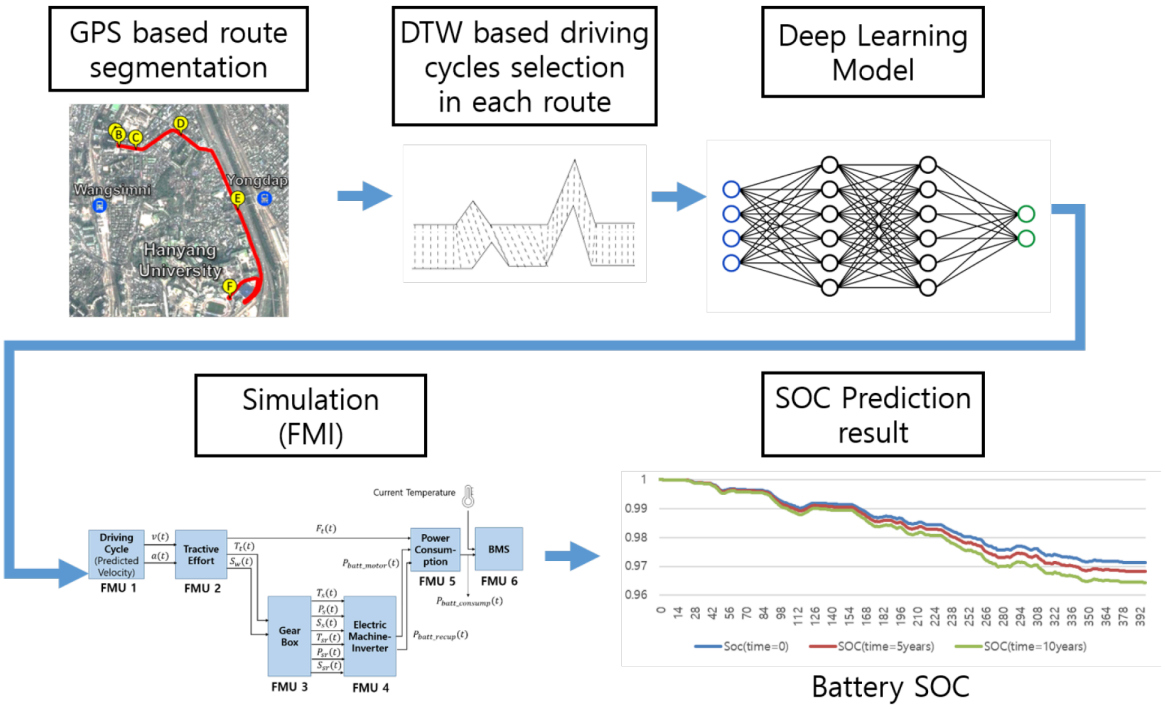


Figure 5: Real Driving Cycle-Based State of Charge Prediction for EV Batteries Using Deep Learning Methods [4]

2. Feature engineering for machine learning enabled early prediction of battery lifetime
3. Machine learning for battery quality classification and lifetime prediction using formation data
4. Real Driving Cycle-Based State of Charge Prediction for EV Batteries Using Deep Learning Methods
5. Deep Dive Analysis 1: Drive Cycle Assessment

Peer Feedback (to be completed next week)

After the submission deadline, you will get randomly assigned 2 blogs from your classmates to review.

Peer review allows you to receive feedback on your writing from multiple peers, which help you see your work more objectively and identify areas where they can revise and improve their writing. It also fosters a sense of community and a safe environment to collaborate and help each other with constructive feedback.

Read the blog posts carefully, paying attention to the content, organization, and style of the writing. Consider the following questions as you review:

- Is the content of the blog post accurate and well-researched?
- Is the structure of the blog post clear and logical?
- Are the ideas in the blog post effectively communicated and easy to understand?
- Is the writing style of the blog post engaging and appropriate for the audience?

Write a review for each of the blog posts, addressing the above questions and any other issues you feel are relevant. Use specific examples from the posts to support your points.