MACHINE LEARNING FOR GREENTECH PROJECT REFLECTION

Moyosoreoluwa Ogunjobi

30156210

Introduction

In this reflection, I want to share my experience working on the GreenTech Innovations Energy Efficiency Prediction project for my ENSF 444 course. Our goal was to use data analysis and predictive modeling to estimate how much energy buildings would need for heating and cooling based on their design features. I'll discuss why I chose this topic, what I personally contributed, the challenges I faced, my presentation experience, and how I think this project will help me in my future career.

Why I Chose This Project

I picked building energy efficiency prediction for several reasons. For starters, I'm really concerned about climate change, and knowing that buildings eat up about 40% of global energy according to the International Energy Agency made this feel urgent and important. What's cool about this area is that even small design tweaks can lead to big energy savings when applied across many buildings.

I was excited about combining data analysis with sustainable engineering to solve real-world problems. Having access to the Energy Efficiency dataset from UCI's repository was perfect since it had clear data points and measurable outcomes that worked well for our analysis. I liked that we were working with actual physical properties of buildings rather than just abstract numbers.

The project also fits with current industry trends toward using data to make better decisions in building design and energy management. I could see practical applications in architectural planning, assessing building upgrades, and informing energy policies, which made this much more appealing than working on something purely theoretical.

What I Contributed

I worked on multiple aspects of this project, from exploring the initial data to simulating how our model could be used in the real world. One of my main contributions was creating a thorough data preparation process. I analyzed the data, created visualizations of the different features and their relationships, and came up with new useful measurements based on my understanding of building physics. I created new features like Volume, Wall-Roof Ratio, Relative Height, and a Compactness-Glazing interaction metric that helped make the relationship between building design and energy use clearer for our analysis.

For the prediction models, I implemented and fine-tuned three different approaches: Decision Tree, Random Forest, and a Neural Network. This wasn't just about using off-the-shelf tools but configuring them specifically for our dataset. I used GridSearchCV to find the best settings for the tree-based models and designed a neural network with dropout layers and early stopping to prevent overfitting. I compared these models systematically using multiple metrics (MAE, MSE, RMSE, and R^2) to make sure we were choosing the best approach rather than just picking one arbitrarily.

One of the more interesting contributions was my analysis of which building features had the biggest impact on heating versus cooling needs. This revealed that while building height mostly affected cooling requirements, surface area had the greatest impact on heating efficiency. These insights went beyond just making predictions and offered practical guidance for designing buildings in different climates.

Finally, I developed a simulation of how our model could be deployed in a real-world setting. I created a Flask-based API for real-time predictions, showing how we could transition from an academic project to a practical application, which is something often overlooked in school projects.

Challenges and What I Learned

The project threw several curveballs my way. The toughest challenge was getting the neural network right. Figuring out the best structure and training approach took lots of trial and error. Despite using techniques to prevent overfitting, the neural network kept underperforming compared to the tree-based models. This was surprising at first since neural networks can theoretically model complex relationships better. But this taught me an important lesson about choosing the right tool for the job: our particular problem (with its distinct clusters and bimodal distributions) actually suited tree-based models better, showing that fancier algorithms don't always mean better results.

Creating meaningful new features was also challenging because it required blending building science knowledge with data analysis techniques. I had to think carefully about which derived measurements would capture real physical relationships while still making sense to people. My feature importance analysis later confirmed these were good choices, as three of my four engineered features turned out to be significant predictors.

Some parts were more straightforward, like implementing the tree-based models and setting up the evaluation framework. The scikit-learn library made this relatively easy, letting me focus more on the conceptual aspects rather than coding details. Creating visualizations with matplotlib and seaborn also went smoothly once I had calculated the right metrics.

I learned several valuable lessons from this project. First, pick models based on your data characteristics and problem structure, not just because they're new or trendy. Second, creating new features based on domain knowledge really pays off, especially for problems with physical foundations. Third, always evaluate models using multiple metrics to get a complete picture of performance. And finally, I gained hands-on experience with the entire process from data preparation to deployment considerations, which gave me practical skills with each step along the way.

Presenting My Work

The demo session was a great learning experience beyond just the technical aspects. Explaining complex details to my classmates and instructors forced me to distill my work into something accessible without sacrificing accuracy, which is a really useful skill in data science. I had to figure out the most important parts of my project and build a story that balanced technical depth with clarity.

During my presentation, the visualizations were particularly effective in showing complex relationships and model comparisons. The feature importance charts generated a lot of interest since they translated the technical model details into practical design insights. This reinforced that it's not just about accuracy but also extracting useful knowledge from your models.

The Q&A session revealed what people were most curious about, particularly why the neural network didn't perform as well as expected and how the feature importance findings could be applied. These questions made me think more deeply about aspects of the project that needed further exploration.

Watching my classmates present was equally valuable, exposing me to different approaches and applications in areas like time series forecasting, computer vision, and natural language processing. I learned about different ways to apply data science techniques and picked up presentation tips by noting which communication strategies worked best with our audience.

The comparison with other projects helped me see both the strengths of my work in model interpretability and practical usefulness, as well as areas where I could improve, like incorporating time-based data or integrating with building information modeling systems.

How This Helps My Career

The skills I developed through this project directly apply to my goal of working in data science within sustainable technology. The technical abilities I gained, from feature engineering to model evaluation and deployment simulation, are fundamental for professional data science roles. My focus on models that not only predict accurately but also provide insights aligns with what companies actually want from data science solutions.

The specific knowledge I gained about building energy efficiency gives me expertise in an area that's becoming increasingly valuable as organizations prioritize sustainability. Being able to translate specific domain problems into data science frameworks is a skill I can apply to many sustainability challenges, from renewable energy to resource conservation.

Beyond technical skills, the project improved my ability to tell stories with data and communicate technical concepts clearly, which is essential for bridging the gap between technical work and business applications. Presenting complex findings to a diverse audience and answering questions on the spot developed my communication flexibility, which is crucial in professional settings.

Looking ahead, I want to expand this work in several ways. First, by including time-based data like weather patterns and building occupancy schedules to improve predictions. Second, by exploring ways to combine different model types to leverage their individual strengths. Finally, by integrating these predictive models with visualization tools or building information software to make them more accessible to non-technical users like architects and building managers.

Conclusion

The GreenTech Innovations Energy Efficiency Prediction project was a thorough application of data science to a significant sustainability challenge. Through this project, I developed technical skills in the complete data science workflow while gaining specific insights into building energy efficiency. The challenges I encountered taught me valuable lessons about model selection, feature creation, and the importance of interpreting results beyond just accuracy numbers.

The presentation experience enhanced my communication skills, while seeing other projects broadened my understanding of data science applications. The knowledge and skills I gained directly support my career goals in sustainable technology, providing both technical foundations and domain expertise for my future work. As sustainability becomes increasingly important across industries, the intersection of data science with environmental challenges represents a growing field where the abilities I developed through this project will be incredibly valuable.