

Predicting Electricity Spot Price

Introduction to Data Science Course Project

Ville Pirsto, Emil Tigerstedt, Ahsan Abbas

October 19, 2024

Contents

1	Introduction	2
2	Project Work Canvas	2
3	Data Collection and Preprocessing	3
4	Exploratory Data Analysis and Visualizations	3
5	Learning to Predict Electricity Spot prices	3
6	Communication of Results	3
7	Summary	3

Introduction

This is the introduction that should contain some pretty general stuff. Below, some initial suggestions are listed.

- What was done?
- Why? Motivation (canvas)
- How?
- Summary of the results

Project Work Canvas

Note: this is just the canvas template, to be replaced with our filled canvas.

MINI PROJECT CANVAS

Title (preliminary):
 Group members:
 Workshop # :

<div style="text-align: center; font-weight: bold; margin-bottom: 5px;">MOTIVATION </div> <p><i>Which is the target group of our mini-project? Who is the end-user?</i></p> <p><i>What are their objectives? What needs do we need to address with our work?</i></p> <p><i>How will they benefit from this proposed solution?</i></p>	<div style="text-align: center; font-weight: bold; margin-bottom: 5px;">DATA COLLECTION </div> <p><i>Which data sources are we planning to use?</i></p> <p><i>Mention database tables, API methods, websites to scrape, etc.</i></p> <p><i>Which is the data management plan?</i></p>	<div style="text-align: center; font-weight: bold; margin-bottom: 5px;">PREPROCESSING </div> <p><i>What are the goals of the preprocessing pipeline?</i></p> <p><i>Give some examples of data preprocessing steps.</i></p> <p><i>What are some possible data cleaning/wrangling methods you're planning to use?</i></p> <p><i>What are some possible data transformations that could be useful?</i></p> <p><i>Any feature engineering necessary?</i></p>	<div style="text-align: center; font-weight: bold; margin-bottom: 5px;">EXPLORATORY DATA ANALYSIS (EDA) </div> <p><i>Look at the data!</i></p> <p><i>What steps are you planning to take towards exploring and understanding better the data you have?</i></p> <p><i>What properties would be meaningful to summarize/visualize in this step?</i></p>	<div style="text-align: center; font-weight: bold; margin-bottom: 5px;">VISUALIZATIONS </div> <p><i>List any meaningful visualizations you are planning to produce that will be useful to the end user?</i></p> <p><i>Are you planning to produce any interactive visualizations?</i></p> <p><i>If so, which types of interactivity might be useful to the end user?</i></p>
<div style="text-align: center; font-weight: bold; margin-bottom: 5px;">LEARNING TASK (focus on problem definition)</div> <p><i>Define the problem setting.</i></p> <p><i>Is this supervised / unsupervised / other...?</i></p> <p><i>Classification / regression / other...?</i></p> <p><i>What are we planning to learn? E.g. What is the target variable / learning outcome?</i></p> <p><i>What variables are we using as input?</i></p>	<div style="text-align: center; font-weight: bold; margin-bottom: 5px;">LEARNING APPROACH (focus on solution implementation)</div> <p><i>Which ML/statistical methods seem more relevant for the defined problem setting and why?</i></p> <p><i>Which evaluation metrics could be relevant?</i></p> <p><i>Is any special treatment relevant regarding how we choose to split the data or how we cross-validate?</i></p>	<div style="text-align: center; font-weight: bold; margin-bottom: 5px;">COMMUNICATION OF RESULTS </div> <p><i>Which type of deliverable will benefit most the end-user? Do we choose to write a blog post, create a website, an app, or other...?</i></p> <p><i>How do we communicate best our results to the predefined target group?</i></p> <p><i>Short description of your interface/workflow (if applicable).</i></p>		<div style="text-align: center; font-weight: bold; margin-bottom: 5px;">DATA PRIVACY AND ETHICAL CONSIDERATIONS (if applicable) </div> <p><i>Are there any fairness constraints that apply to our proposed pipeline?</i></p> <p><i>Is there a need to ask for consent during the data collection process?</i></p> <p><i>Is there a need for data pseudonymization/anonymization?</i></p> <p><i>Any other privacy considerations that come to mind?</i></p>
<div style="text-align: center; font-weight: bold; margin-bottom: 5px;">ADDED VALUE </div> <p><i>Is there a possibility for added value from the data we're planning to use?</i></p> <p><i>What is the added value?</i></p> <p><i>How are predictions turned into added value for the end-user?</i></p>			<div style="text-align: center; font-weight: bold; margin-bottom: 5px;">LEGEND</div> <p>WEEK 1: Data collection/preprocessing</p> <p>WEEK 2: EDA & visualizations</p> <p>WEEKS 3-4: Machine/deep learning</p> <p>WEEK 5: Fairness & data privacy</p>	

Figure 1: Canvas used for project planning.

Data Collection and Preprocessing

- Where is data used from?
- How is data stored?
- What kind of data is used?
- How is data accessed?
- What kind of preprocessing is done?

Exploratory Data Analysis and Visualizations

- What kind of EDA was done?
- How was the data visualized?
- What kind of findings were obtained?

Learning to Predict Electricity Spot prices

- What kind of approaches were tried?
- What kind of observations were made during this process?
- What approach was used in the end? Why? How did we end up to it?

Communication of Results

- Webpage, how?
- What kind of user interface is used?
- What does the webpage do? What can the user do?

Summary

- Predicting electricity spot prices is difficult
- Nevertheless, a decent method was found for the used input variables
- We learned alot, maybe list here someof the thigs that we discussed
- What could be done better / future continuation topics