convista:

HS Ansbach Summer Term 2023 – Advanced AI

Pracitcal Assignments – PV Power Surplus Prediction – 26.05.2023

1_ General Information

There will be marked assignments you will need to complete to pass this course.

Make sure that you are a member of the correct group in our Moodle course.

All the information regarding the groups will be available in Moodle.

Please upload your material to Moodle until July 21st, 2023:

- Python notebook(s) of the code (make sure it is reproducible)
- Optional: group presentation from June 30th, 2023
- Optional: Other explanatory documents (it may be easier for you to work on several steps in other document-types like Text-Documents, Tables, Slides, …)

We require 10 pages of material per participant (remember that you can submit code, pictures, text, …).

Please ensure to denote/mark who contributed what part in order to allow for an individual evaluation.

If you have questions regarding the project please use the forum in the Moodle course!

2_ Task Description

Data: "SesKiPvEletric.csv"

Historical data of hourly PV output from previous years with explanatory variables (radiation, humidity, temperature). In addition the power taken from the grid and fed into the grid is given.
Note: The data is given in german numeric representation, i.e. "," is used as decimal-separator instead of "." (1,34 = 1.34). For the columns "El. Energy Heating/Cooling", "El. Energy house", "PV generation", "From grid" and "To grid" you have two columns

each. One describes the cumulative value one the measurement for the current hour. Example: For the value on "01.01.2020 01:00:00" "PV power generated 280; cumulative" gives you the total power generated over all time until "01.01.2020 01:00:00", "PV power generated 280" gives you the power generated between "01.01.2020 00:00:00" and "01.01.2020 01:00:00".

General Task: Describe and analyze the data at hand in a suitable way. Come up with a Deep-Learning-Model that predicts the surplus (or defecit) in power generated from the PV at the Campus Feuchtwangen compared to the consumption. Use other covariates as you see fit. Use at least one Deep Learning model discussed in the lecture, a Multivariate-LSTM is recommended.
Evaluate the models performance against at least one other model, i.e. a statistical model (ARIMA), univariate-LSTM, feedforward NN or a Machine Learning model.
Optional: You can elaborate on the ways such a prediction can be used in the context of smart energy management; can you outline potential benefits of battery capacity, electric vehicles? What would be a sufficient battery capacity to maximize benefits?
If you want a more step-by-step approach you can follow the tasks below:

Tasks:

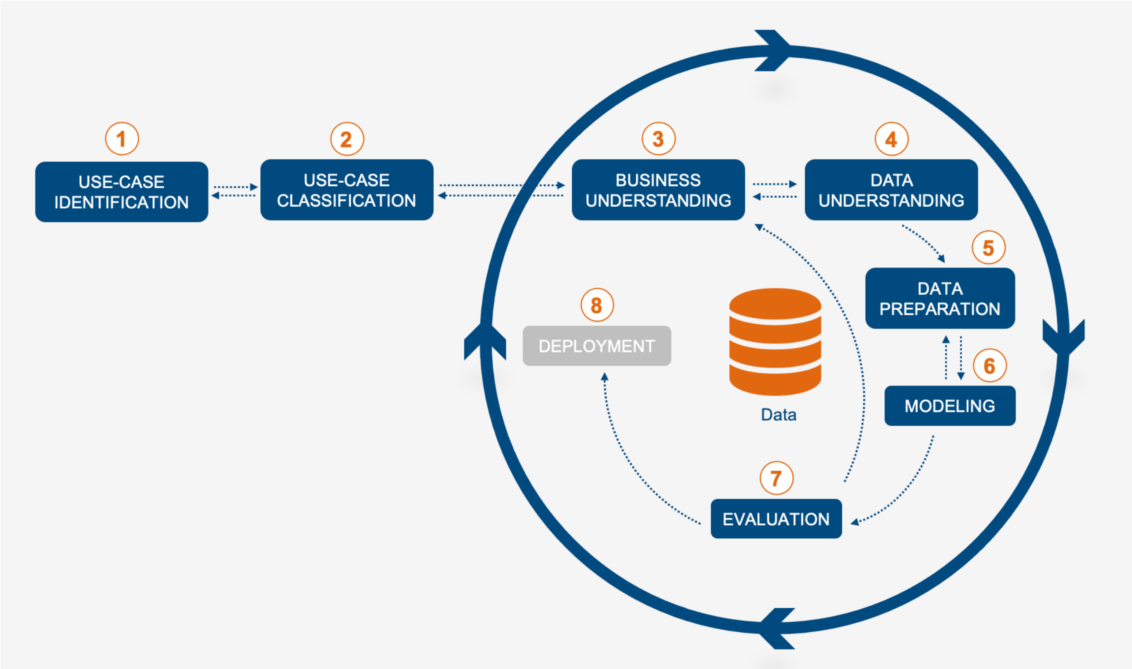
1. Outline the Business Value of Use-Case. Outline the criteria a model would have to fulfill in order to add value. (see Phase 3 in CRISP-DM Model)
2. Describe the data set at hand in-depth (see Phase 4 in CRISP-DM Model)
    a. Derive the target variable; you can use "To grid"-"From grid".
    b. Name and describe all target variable in the data
        i. Basic Statistical Analysis of the target variable (distribution, mean, variance, interpretation)
    c. Name and describe all the explanatory covariates in the data
        i. Basic Statistical Analysis of the covariates (distribution, mean, variance, interpretation)
    d. Give an overview over the data set as a whole
        i. Illustrate and interpret the correlation between (some of) the covariates
    e. Describe suspicions values like outliers, missing values ("na") or obvious patterns in the data if you encounter any. Give some possible reasoning for the observations and assess how to deal with those.
3. Preprocess the data to make it ready for use in Machine Learning (see Phase 5 in CRISP-DM Model)
    a. Load the dataset into a Python-Notebook

   b. If necessary: Resolve the issues encountered in 2.e.
   c. Rescale/normalize some of the parameters if that makes sense.
   d. Optional: Derive additional paramters from the data. For example one could derive the time of the year/month/time-of-day from a timestamp and use it in the further steps.
   e. Split the data into train and test-data (possibly train, dev and test data if you need to) in a suitable way.
4. Train a DL-model for a relevant architecture encountered in the course (Recommendation: Multivariate-LSTM) in order to predict the power consumption in the next hour. (see Phase 6 in CRISP-DM Model).
   a. Describe why you want to use the model(s) of your choice for this specific use-case.
   b. Train a baseline model first in order to compare the performance of your main models. Suitable baseline-models: univariate-LSTM, ANN, general Time-Series Model (ARMA, ARIMA, GARCH, …), general Machine Learning Model.
   c. Train a DL-model using the preprocessed data from 3.
      i. Recommended: Use techniques (grid-search, cross-validation, regularization, …) to improve the model performance
5. Evaluate the model trained in 4 on the evaluation criteria defined in 1. Use the feasible data set for the evaluation (usually your test set). Interpret the results. Write a project conclusion and outlook of the next steps you would recommend. (see Phase 7+8 in CRISP-DM Model)
   a. Opional: You can elaborate on the ways such a prediction can be used in the context of smart energy management; can you outline potential benefits of battery capacity, electric vehicles? What would be a sufficient battery capacity to maximize benefits?

3_ Proposed Project Structure

It is recommended to structure the assignments using the well-known Crisp-DM model.

We have extended this model by two additional phases (1 and 2) to illustrate the challenges encountered when identifying and evaluating a potential Use-Case.



An in-depth explanation of the classical Crisp-DM (phases 3-8 in the graphic above) can be found in many places on the internet (good example: With, Hipp: CRISP-DM: Towards a Standard Process Model for Data Mining, 2000).

Let us briefly describe what you could do in the different phases of the model and how you can structure your assignment:

1) *Use-Case Identification:* You will need to think about potential Use-Cases in your domain of Smart Energy Systems: Where do we have (a lot of) data? What are interesting problems in our field that one could evaluate in a data-driven way.

2) *Use-Case Classification:* Evaluate the Use-Cases identified in 1) and come up with the most promising one. Possible criteria: Amount, quality and availability of data. Business value in solving the Use-Case. After this phase you would have identified one Use-Case for applying DL. In the following points

3) *Business Understanding:* Understand and outline the objectives and requirements of the project. Assess the current situation and how using Machine Learning/Deep Learning can be of benefit. Take risks into account as well. Define your goals and what you wish to achieve. This has to be taken into account in the “Evaluation” phase at the end of the cycle. The output is a project plan where you select technologies you want to use (which forms of DL models are to be used) and define a plan for the ensu-ing phases.

4) *Data Understanding:* Collect the data set. Describe and explore the data (what are the explanatory variables, for supervised learning: what is the target variable (label)). Possible approach: Use visualization tools, correlation analysis, PCA, describe do-mains of the covariates, … Verify data quality and document issues like anomalous,

unplausible or erroneous data. You should be able to show that you understand the data and its challenges.

5) *Data Preparation:* Select the data you want to use. Clean the data is very important (“garbage in - garbage out”). Derive other interesting variables (feature engineering). You possibly have to integrate data from different sources. Choose a feasible ma-chine-readable format for the covariates. Look into standardizing the data which is generally a good idea. Split the data into train, validation and test sets.

6) *Modeling:* Now you can start actually training DL models. Sometimes it is a good idea to first establish a baseline model with more simplistic approaches from classical Ma-chine Learning. Select a DL-algorithm (or multiple) that you want to try. Then build your model. This is just one of the many steps, often it is a rather simple one. You will typi-cally revisit this phase after the “Evaluation”-phase: Try different hyperparameter combinations. Maybe you implement a grid-search algorithm in order to find better hy-perparameters in an analytical way.

7) *Evaluation:* Evaluate you model(s) with the previously defined criteria (Business Un-derstanding). This is a good point to take a step back and look at the process: Can you improve on any of the previous steps? Determine the next steps: Is the perfor-mance good enough? If not: Maybe you revisit the previous phases with your new and improved understanding of the subject. At the end of this phase you should be able to come up with a wholistic evaluation: Which is the most useful model? What can you achieve with the model? What are its possible limitations?

8) *Deployment:* Outline the next steps needed to make the model “production ready”? You should produce some kind of final report of what you did during the previous phases and review the project.