CHAIR OF PRODUCTION MANAGEMENT
Prof. Dr. Raik Stolletz
http://stolletz.bwl.uni-mannheim.de

UNIVERSITY OF MANNHEIM

Assignment 4 - Supervised learning for data driven tomato yield prediction and control of greenhouses

OPM 562 - Business Analytics: Applications of Artificial Intelligence for Data-Driven Decision Making

Deliverables:

- A Jupyter notebook with explanations and commented code cells used for your analysis
- Trained Neural Network as **h5** file
- A presentation in PDF format (maximum of 15 content slides)
- A recording of your presentation via PowerPoint (15 - 20 min). Here is a "How to" guide

Deadline: May 26th, 2020, 23:59

Assignment Type: Group assignment

Total Points: 40

Introduction

Kelby Company Ltd. is a multinational agriculture company with greenhouses in Spain and the Netherlands. In an effort to account for the trend of locally grown fruits and vegetables supported by many supermarket chains, Kelby Company Ltd. plans to establish a new large site in Germany. The company needs to decide on how much space in the greenhouses should be devoted to the different vegetables and how to control key parameters of how they operate the greenhouses. Advanced technology in modern greenhouses (see Figure 1) allows for computer climate control that permits a year-around production with up to six harvests per year.

Figure 1: Modern greenhouse used by Kelby Company Ltd.

The appointed branch director in charge of tomato production is Ms. Holly Kelby. She heard about the predictive power of machine learning approaches and wants to utilize it for the control of the new site. With a background in agriculture studies, she is an expert in the cultivation of vegetables and fruits. Hence, she knows that the key parameters that determine the obtainable yield per square meter are: irrigation, pesticide treatment, temperature, and exposure to sunlight.

The irrigation applied per day per unit area is one of the most important parameters driving the growth of the plants. Watering the tomatoes insufficiently may result in low yield rates. Overwatering on the other hand, not only incurs higher costs but also may result in disease for the plants.

To grow healthy tomatoes, it is important to apply proper pesticides. Different countries have different regulations allowing different kinds of pesticides. The type of pesticide used is the only structural difference between the existing sites. Usually, tomato growers choose the most powerful kind of pesticide allowed in their country. The most powerful pesticides allowed in the Netherlands and Spain are Envidum and Glyfanac, respectively. Both Envidum and Glyfanac are allowed in Germany and available for approximately the same price.

Temperature is another factor that plays a key role in the production of high-quality tomatoes. To maintain the temperature at the desired level, modern greenhouses are equipped with smart heating and cooling systems. Deviations from the optimal temperature may result in lower yield as well as extra heating/cooling costs.

Light is essential to the process of flowering and fruiting in tomato plant growth. The most important source of light in the Kelby greenhouses is the sunlight. The available sunlight varies throughout the year depending on the season.

Ms. Kelby is aware that the training of machine learning models requires meaningful input data. Hence, she asked an intern to collect data from the existing sites in the Netherlands and Spain.

Dataset

The collected data set is provided as “data.csv” and consists of 1894 rows, each of which contains information about one cycle of tomato harvesting for one modular greenhouse. In the Netherlands and Spain, these greenhouse modules are controlled individually by an operator. Each cycle coded as C1 to C6 indicates a period from planting to harvesting the tomatoes at different times during the year. A cycle encompasses approximately 60 days. The data set includes the following columns:

- **cycle**
- **country**
- **irrigation:** Average daily irrigation applied ($\frac{\mathrm{L}}{\mathrm{m}^2\,\mathrm{d}}$)
- **sun:** Average number of sunny minutes per day throughout the cycle ($\frac{\mathrm{min}}{\mathrm{d}}$)
- **temperature:** Average temperature during the cycle inside the greenhouse ($^\circ\mathrm{C}$)
- **yield:** Total marketable tomatoes produced in the cycle per unit area ($\frac{\mathrm{kg}}{\mathrm{m}^2}$)

Tasks

Use supervised machine learning techniques to support the planning of the new site in Germany. To help Ms. Kelby, please complete the following tasks.

Task 1 Data Preparation (3 Points)

Read and prepare the data such that it can be used for training, validation, and testing of supervised learning algorithms.

Task 2 Neural Network (25 points)

Create and train a neural network which predicts the yield of tomatoes per m^2 given the features:

- **cycle**
- **pesticide:** Type of pesticide applied depends on country
- **irrigation:** Average daily irrigation per unit area ($\frac{\mathrm{L}}{\mathrm{m}^2\,\mathrm{d}}$)
- **sun:** Average number of sunny minutes per day throughout the cycle ($\frac{\mathrm{min}}{\mathrm{d}}$)
- **temperature:** Average temperature inside the greenhouse during the cycle ($°\mathrm{C}$)

Explain the process of how you selected the architecture of the neural network, the activation functions, as well as the loss function. Justify the selection of hyper-parameters if applicable and describe how you ensured that the trained model does not suffer from overfitting.

Task 3 Decision making (8 Points)

The company has a contract with several supermarket chains in Germany for six months corresponding to three cycles starting with cycle C2. The amount of tomatoes that have to be harvested per cycle, the sunlight in each cycle, and average outside temperature are given in Table 1.

Table 1: Sunlight, outside temperature, and demand per cycle for the German site

cycle	sunlight (min/d)	temperature outside ($°\mathrm{C}$)	demand (kg)
C2	278.5	15	300,000
C3	426.0	22	500,000
C4	417.0	25	580,000

The penalty costs of under-fulfilling the demand are 1 EUR/kg. In case the produced tomatoes exceed the demand, the excessive tomatoes can always be donated without additional costs.

To achieve the desired temperature inside the greenhouse heating or cooling is used. Consider costs of 3600 $\mathrm{EUR}/\mathrm{cycle}$ for each $°\mathrm{C}$ that the inside temperature is changed compared to the outside temperature. The new greenhouses allow for temperature control in steps of 1 $°\mathrm{C}$.

The irrigation is another key decision by the greenhouse operator. The costs are 0.021 $\mathrm{EUR}/\mathrm{m}^3$. The average irrigation applied can be varied in steps of 0.5 $\frac{\mathrm{L}}{\mathrm{m}^2\,\mathrm{d}}$.

Moreover, there are variable costs for using an area in the greenhouse for tomato production, which include costs for the depreciation, and labor cost. These amount to 20 $\mathrm{EUR/m^2/cycle}$. Note that the maximum area that can be used for tomatoes at the German site in each cycle is limited to 50000 m^2. The greenhouse modules used at the German site are of size 1000 m^2. Assume that for efficiency reasons always complete modules are used for one type of vegetable.

The company plans to use central control of all greenhouse modules. Determine the optimal pesticide, applied irrigation, chosen inside temperature, and area devoted to tomato production that minimize the sum of all relevant costs. You can determine the optimal decisions by systematically iterating over all potential decisions while utilizing the model developed in Task 2 to predict the total costs.

Clearly state your recommendations for the decisions and the resulting total costs.

Task 4 Sensitivity Analysis (4 Points)

Ms. Kelby wants to know how sensitive the results are with respect to the contracted demand in case the customers change their orders. Conduct a sensitivity analysis on the demand by varying the demands compared to the data provided in Task 3 from -20% to +20% (for all three cycles) in steps of 10%. Report the resulting optimal decisions and costs. Provide the company with insights on how the costs and decisions will change if the demand changes. Use visualizations and tables to support your analysis.