* Anpruch an die predictive Analyse 🡪 Vorhersagen
* Aufzählen, was wir verwenden (Tree Classifier, RNN, Julians und Kims Sachen)
* Parameter erklären, begründen
* Evaluation
* Vergleich 🡪 Ranking

ARIMA-Modell:

* Anpruch an die deskriptive Analyse 🡪 Möglichst guter fit ohne gute zukunftsprognosen
* Aufzählen was wir alles verwenden (ARIMA, Tree-Model)
* Grid search 🡪 Beste parameter
* Bester fit
* Aussage erklären! Hidden systematologies? Wie ein Model aussehen würde, wenn es welche gäbe.
* What trading suggestions would you derive?

Descriptive Analysis

After the basic understanding of the data and the additional preparation the first methods can be applied in order to accomplish the first goal of this analysis: Finding hidden pattern in the data. The methods used for this purpose should be applicable to time series data and fit the data well. Furthermore, the most important feature is interpretability of the model since the second goal is to derive potential trading suggestions and these can only be deduced if the model is transparent in which and how the respective attributes are causing a potential hidden pattern.

The first two features are provided by statistical ARMA based models. The presumption for ARMA models is that the time series is stationary. That means that mean and standard deviation ought to be stable over time. ARMA models are divided into an “Auto Regressive” (AR) and a “Moving Average” (MA) part. In order to describe a particular timeslot t the auto regressive part represents the influence of the last p timeslots and the moving average part contributes of the last q timeslots. Therefore the standard parameter for ARMA are p,q which are extended with several others depending on different modifications of ARMA such as ARIMA (Auto Regressive Integrated Moving Average), where I is the parameter that describes the number of differencing in order to make the time series stationary. Other modifications of ARMA models are SARIMA, which considers seasonality or ARIMAX, which considers exogenous variables. The biggest drawback of ARMA type models is the difficulty of interpretation, which is one of our criteria. Thus ARMA models are not used in this part, but in the predictive part of our further analysis. Other opportunities especially for pattern recognition in time series are seasonality and trend. Both meets all the required criteria. A visual analysis shows that trend can be obviously excluded, but seasonality more difficult to assess. [PLOT PLAIN TIMESERIES] Hence a ACF-Plot shows which previous timeslots are most correlated with a specific timeslot. [PLOT ACF] The plot states out that there is a high autocorrelation between a timeslot t and the timeslot t-96, which is the timeslot 24 hours before t. Another peak can be observed at 672, which describes a high autocorrelation between t and the timeslot one week earlier. (TODO: ADDING AN EVALUATION FOR A SIMPLE MODEL THAT ONLY CONSIDERS SEASONALITY BASED ON ITS AUTOCORRELATION- t = 0.6\*t-96 + 0.4\*t-672 🡪 evaluation evt auch über decomposition/residual)

(TODO: weitere modelle die speziell auf time series daten spezialisiert? – ordinale ts analyse?)

Apart from models which are designed especially for time series data there are other models that can be taken into account such as a decision tree model. The major advantage of a tree model is its interpretability. In order to find the best model parameter with the least number of leafs that fits the data well, a grid search is implemented for the model. In addition, a time series capable cross validation is implemented to verify the accuracy, which is used to measure how well the model fits the data. The resulting Tree shows that the model uses only three variables to identify whether the “price premium” is positive or negative: “dailySunnyHoursAvg”, “numberRainyDays” and “participants”. [tree model plot]

The accuracy is 78.89% (and the confusion matrix is shown in the following.)

[[38012 6582]

[ 8068 16746]]

[TODO: trading suggestions einfügen wenn Tree ergebnis geklärt -> normalisiert anderes ergebnis ohne]

Predictive Analysis

The predictive analysis is in contrast to the descriptive part not focused on interpretability and goodness of fit. The only dimension of interest is the predictive power, which is represented through accuracy throughout this project. In order to simplify the interpretation of the performance through the accuracy a baseline is calculated first. Therefore, a simple “ZeroR” classifier is introduced, which predicts only the class which has the most observations in the dataset. This technique simulates random predictions. With regard to our dataset the baseline is: . Above mentioned the tree classifier can also be used for out-of-sample predictions and with reference to its accuracy 78.89% it can be stated out that it is around 14% better than the defined baseline.

As prefigured in the descriptive part, there is also an ARMA type model mentioned for the predictive task. In particular an ARMAX model with [ARMAX nochmal genau anschauen wegen predictions so mies…]