**INTRODUCTION**

* Introduce research/science of mood predictions related to digital application usage
* Introduce temporal data mining and comparison of classical (parametric) temporal models (ARIMA, RNN) and supervised (non-parametric) machine learning models (RF, SVM)
* Indicate objective and results of this assignment

**PRE-PROCESS THE DATASET**

Treat outliers and strange values – EDA

Dropping of values before pivot – do EDA on raw and mature dataset

Shortly explain procedure of what have been done

Give descriptive statistics / plot information about dataset

Explain rationales of choices that have been made

* Length of lookback/aggregation window (number of days defined by looking at autocorrelation of mood?)
* Feature creation – what features are created and why?
* Structure of instance based dataset
* Treatment of missing values – Interpolation / imputation
* Feature selection method (correlation, forward/backward feature selection)
* Check AC/PAC – 27,29,31

**LEARN USING THE DATASET**

Why single model for all patients?

Choice of ML techniques – why RF and SVM? (based on instance based dataset)

Algorithm that is able to cope with temporal data

Implementation of a benchmark model

Performance measures – explain choice of measure and state formula

Compare performance based on measure (create a table)

Plot prediction results

Plot feature importance results

Use graphs to illustrate the performance in an insightful way

**EVALUATE AND REFLECT ON YOUR RESULTS**

Analyse the results in detail both using a more statistical view and by means of your interpretation. Argue what the pros and cons of the different approaches are.

ARIMA

<https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/>

<https://www.projectpro.io/article/how-to-build-arima-model-in-python/544>

**Tutorial 1 🡪 temporal data mining**

Brimag\_Exam (canvas channel)

Google Scholar

Time data processing in Pandas

Test, Validate, Test

Export figures as pdf

Make GitLab ready

Task1: Deal with preparing data for time-series usage

Data has been used, temporal data mining, aggregation methods, target window

Instance based dataset

Historical data can be used up to from different length

Baseline model, mood of the next day is mood of yesterday

Carefully read the instruction for academic writing

Explain rationales how solutions are developed

No right or wrong, just explain

<https://github.com/Dantesean/dmt-1>

<https://github.com/Dantesean/dmt-1/blob/master/DMT%20A1%20Preprocess.ipynb>

<https://github.com/Dantesean/dmt-1/blob/master/cleaning.ipynb>

<https://github.com/Dantesean/dmt-1/blob/master/DMT%20Assignment%201_Dante.ipynb>

<https://medium.com/data-science-at-microsoft/introduction-to-feature-engineering-for-time-series-forecasting-620aa55fcab0>

<https://www.bi4all.pt/en/news/en-blog/supervised-machine-learning-in-time-series-forecasting/>