Advanced Assignment Approach

|  |  |
| --- | --- |
| Team nr: | **64** |
| Names: | **Arran Lyon Viktor van der Valk Walter Vianen** |

# General remarks

Please fill in the form below. Be sure to carefully think about your approach and look for scientific literature to support it. Try to be clear and **concise** in your explanations. If you are unsure about important parts of your approach (in particular about the preprocessing), then it is best to make an appointment. Please send this form or your request for an appointment to [karel.dewit@student.uva.nl](mailto:karel.dewit@student.uva.nl?subject=Approach%20of%20group%20nr:%20X).

## Dataset and Preprocessing

The raw dataset is not yet usable. You need to transform the data focusing on two aspects. First, the data should be in a format which can be used by the algorithms you selected. Since you will use a temporal and a non-temporal algorithm, you might have to create two datasets with very different structures. Second, you should create attributes to enhance the performance of your algorithm. For example, a running average of the mood over the last X days.

|  |  |
| --- | --- |
| Task | Approach |
| How will you create a *usable* instance-based dataset? | For each patient create a table like this:   |  |  |  |  | | --- | --- | --- | --- | | **Day** | **Mood** | **Screen** | **…** | | 1 |  |  | … | | 2 |  |  | … | | 3 |  |  | … | | … | … | … | … |   In which all values for the variables are averages, or totals of that day. Then clean up the tables. Divide the data into test and training data (e.g. last 10% of days per patient is test data). Use this data to make intervals of size (to be determined) to make a table like this:   |  |  |  |  | | --- | --- | --- | --- | | **Period** | **Mood** | **Screen** | **…** | | [1, n] |  |  | … | | [2, n+1] |  |  | … | | [3, n+2] |  |  | … | | … | … | … | … |   In which all values for the variables are (possibly weighted in case of temporal model) averages of the variable values of the days before. |
| What attributes will you create to enhance performance? | The tables above, in the process of organizing the data, will contain the following measures:   |  |  | | --- | --- | | **Variable** | **Metric** | | Mood | average | | circumplex.arousal | average | | circumplex.valence | average | | Activity | average | | Screen | total time | | Call | total amount | | Sms | total amount | | appCat.builtin | total time | | appCat.communication | total time | | appCat.entertainment | total time | | appCat.finance | total time | | appCat.game | total time | | appCat.office | total time | | appCat.other | total time | | appCat.social | total time | | appCat.travel | total time | | appCat.unknown | total time | | appCat.utilities | total time | | appCat.weather | total time | |
| What scientific literature supports these choices? (provide URLs) |  |
| What is your target variable? | *The average mood over the entire next day.* |
| Other remarks? | We do some feature selection based on:   * A minimum sample correlation coefficient:   And then for the feature to be selected:  Where some threshold value we consider the minimum for sufficient correlation.   * Possibly some t-testing. |

## Learn using the dataset

In the next step, you are going to use your dataset to create a predictive model. You are asked to use a temporal and non-temporal algorithm. Think about how you want to evaluate the results and (statistically) compare them to the benchmark. Since the lecture on evaluation is on the 19th, it is ok if you’re not very sure about the performance metric and evaluation approach.

|  |  |
| --- | --- |
| Task | Approach |
| What non-temporal algorithm will you use? | Random Force Classifier |
| What temporal algorithm will you use? | VAR / ARIMA |
| What performance metric will you use? | [OK if you are unsure about this] |
| How will you evaluate your algorithms? | [OK if you are unsure about this] |
| Other remarks? | We’re unsure about the specific definition of temporal vs. non-temporal algorithm. In the assignment description there is no mention of this. What would a non-temporal algorithm take as input vs. a temporal algorithm? |