

BUSINESS ANALYTICS: EXERCISE SHEET No.5 (WiSe2018-19)

**22ND OF NOVEMBER (DUE 29TH OF NOVEMBER
AT 23:59:59)**

Dr. Josif Grabocka, Rafael Rego Drumond
HiWi: Kiran Madhusudhanan
Information Systems and Machine Learning Lab
University of Hildesheim

QUESTION 10 - TIME-SERIES CLASSIFICATION - 10 POINTS

(Type C) Consider a SID problem, where from audio clip you must identify the speaker ("who is speaking?"). You are given a data-set with several audio clips where we have Z different speakers and each speaker has recorded a number X of audio clips saying a set of Z passwords (each password is related to each speaker) multiple times. In other words, each speaker i is associated with a password j , and, on our data-set, every speaker records multiple audio-clips for every existing password (including the ones from other speakers).

In order to solve our problem there are two sub-tasks to be completed:

- Identify the speaker ID.
- Identify if the password is the correct one for that speaker ID.

How would you solve each sub-task separately using shapelets? How would it be possible to solve both tasks at once using shapelets in a single classification? To answer these questions formalize each problem and explain how would you model each task. Also explain one advantage and one downside for each approach.

QUESTION 11 - TSC WITH PAA AND SAX 10 POINTS

(Type B) Consider the two Time series:

·
·

T_1	T_2
0.84	1.00
0.98	0.71
0.54	0.00
-0.21	0.00
-0.84	0.00
-0.98	0.00
-0.54	0.00
0.21	0.71
0.84	1.00
0.98	0.71
0.54	0.00
-0.21	0.00
-0.84	0.00
-0.98	0.00
-0.54	0.00

a) Plot both time series. Consider that our break points are $\beta_1 = -0.25, \beta_2 = 0.25$ and length $L=2$ perform PAA and then SAX (with words of size 2).

b) Using a nearest neighbor classifier, predict the class for T_1 and T_2 using the SAX histogram for each class in the table below. (code allowed for this part)

aa	ab	ac	ba	bb	bc	ca	cb	cc	Class
3	2	2	2	3	2	2	2	3	0
0	0	0	0	4	1	0	1	2	1

Consider $c > b > a$.

WARNING!

If we detect **Plagiarism** on your solution, you will receive no points for it. If a second plagiarism attempt is detected, you might fail the class or be expelled from your program.

You are allowed to discuss solutions, but if you work on a group, you must indicate on your sheet with whom are you working with.

Group submissions earn 0 points.

BONUS POINTS!

Submission grades represent 100% of this bonus. For each 11% of your total grade you will earn 0.5

bonus points on the final exam up to a maximum of 4 points.

HOW TO SUBMIT?

Inside the Samelzonplatz Campus, there is a post-box cluster to the right corridor from the entrance. Look for the "Business Analytics Post-Box" (we will set it up on Monday). For security you must also submit on our learnweb page: <https://www.uni-hildesheim.de/learnweb2018/course/search.php?search=3108> (password: 3108). If you submit your physical version you will get it back with correction markings. Submitting on learnweb will guarantee that you will receive your grade (in case you forget to post it or it goes missing).

Keep in mind that your solution sheets are important documents, so make sure they are clean and organized. Non-readable sheets will not be graded.

WRITE YOUR NAME, ID, AND TUTORIAL GROUP.

REMEMBER TO KEEP A COPY OF YOUR SOLUTION FOR YOURSELF!