

# Term Project - Deep Learning

## (1) Dataset description:

A tabular time-series dataset (csv file) with **290 monthly** records (from 1997-2021). Each monthly record (row) contains four features, named as **date**, **rate**, **ems**, **pkh**.

## (2) The ML model:

Implement a recurrent neural network by **Pytorch** based on **RNNs or LSTM**. The neural network **must be** trained with three columns **rate**, **ems**, **pkh** as **inputs** and column **pkh** as an **output**. The column **date** shows only the time order of the monthly data.

## (3) Defining the training and test datasets:

Select **the first 80%** (the first 232 rows) , and **the rest of 20%** (last 58 rows) of the csv file for the training dataset and test dataset respectively. The two datasets at this phase must not be shuffled to preserve the order of the monthly data.

## (4) Creating training and test samples:

Each sample in the train dataset should be a sequence with the **length of 231 elements**, and each sample in the test dataset should be a sequence with **the length of 57 elements**. To make the lengths of all samples equal, a zero padding technique is used at the beginning of each sample. The targets (prediction values) for both training and test samples are selected from **pkh** column while creating the samples as the following:

Initially, a **sliding time-window of length  $L=10$**  is defined which scans the data row by row and moves **1 step every time when a new sample is created**. In this way, for two consecutive samples  $X_i$  and  $X_{i+1}$ , the sample  $X_{i+1}$  contains 1 more row of data compared to that of  $X_i$  and, therefore,  $X_{i+1}$  has 1 less row of zero padding. The target value (pkh) of each sample is selected from **the first row that is not within the sliding window** and located at **the right side (head) of the sliding window**.

$X_1$  (training sample)

rate	ems	pkh	$\leftarrow t_1$
0	0	0	
0	0	0	
.	.	.	
.	.	.	
.	.	.	
0	0	0	
3.45	22.93	790	
2.15	15.46	539	
1.89	12.61	535	
2.03	12.84	523	
2.25	15.38	552	
2.2	13.43	589	
2.19	11.59	592	
2.49	15.1	894	
2.88	14.87	861	
3.07	14.9	645	$\leftarrow t_{231}$

The 1st 221 rows  
of  $X_1$  are set with  
zero padding

$r_1, \dots, r_{10}$  from training dataset

target of  $X_1$

NA    NA    **642**    The value of pkh column at  $r_{11}$

$X_2$  (training sample):  
features : [the 1st 220 rows with zero padding +  $r_1, \dots, r_{11}$  ]  
target : The value of column pkh at  $r_{12}$

·  
·  
·

$X_n$  (last training sample):  
features : [ $r_1, \dots, r_{231}$ ] //no zero padding  
target : The value of column pkh at  $r_{232}$

Create the test samples by the same technique using the test dataset.

#### (4) Documentation and Submission

- Project documentation ( \*.doc):

- 1) Introduction
- 2) Description of the ML model and its architecture.
- 3) Report results on both training and test (losses and accuracies), and training/test graphs.  
The results must be depicted on both the .ipynb program file and also in the form of screenshots in the .doc file.
- 4) **Implement** an improvement of the prediction outcomes **based on modifications to the model's architecture**. Describe the modifications and report the new outcomes as stated in sections (2), (3).
- 5) Interpretation and discussion of all results.
- 6) Conclusion
- 7) References
- 8) The tasks that each member of the group performed during the project activities.

- The program files (\*.ipynb) must have necessary comments.

- Submission: **two \*.ipynb files** (for the initial ML model and its modified model) and **one \*.doc file** and place the files in a ZIP file, and submit the ZIP file **ONLY** on BB.