Pedro Francisco Sousa e Silva PG50689, Tiago Aguiar Martins PG50775 e Dong Xuyong PG50343 (TP2 Grupo 5).

1 Introdução

Within the subject of Artificial Intelligence Techniques in Forecasting and Optimization in Business Systems, we were proposed a project aimed at using forecasting and optimization techniques in a real mind problem, in this case the distribution of drinks in a company.

Throughout this document we will first analyze the beverage sales data, then we will apply forecasting techniques to predict the number of sales of each beverage, this forecast will be divided by a univariate forecast and a multivariate forecast, for both will be used the knowledge and practices acquired in class to predict successfully. Finally, we have Optimization, which consists of finding the best values in order to maximize the sales value of each drink.

2 Dataset and Requirements

For this project, we were provided with an excel file called "bebidas.xlsx", within which are the daily sales records of each of the two beverages made available by the company in question, within that excel file there are still other relevant data, which will be detailed afterwards.

2.1 Parametros do dataset

In the following image we have a print of the columns of the dataset mentioned above:

	A	B	C	D	E	F	G
1	DATA	DIA_SEMAN/	PRECIPITACA	TEMP_MAX	STELLA	BUD	
2	01/02/2019	4	6,8	30,1	53	71	
3	01/03/2019	5	0	32,9	106	235	
4	01/04/2019	6	14,2	31,8	218	42	
5	01/05/2019	7	3	27,7	180	110	
6	01/06/2019	1	0,6	29	69	15	
7	01/07/2019	2	0	31,6	18	8	
8	01/08/2019	3	0	33,2	61	10	
9	01/09/2019	4	0	31,1	38	6	
10	01/10/2019	5	0	33,2	545	26	

Figure 1: Project Dataset

The following dataset is compose of 6 columns, they being:

•DATA: This column represents the date the records are from;

•DIA_SEMANA: This column represents the day of the week, where 1 is Sunday, 2 is Monday, 3 is Tuesday, 4 is Wednesday, 5 is Thursday, 6 is Friday and 7 is Saturday;

•PRECIPITACAO: This column represents the total of precipitation in mm in that day;

•TEMP_MAX: This column represents the daily maximum temperature in Celcius from that day;

•STELLA: This column represents the number of STELLA drinks that where sold in that day;

•BUD: This column represents the number of BUD drinks that where sold in that day.

2.2 Graficos

semanasxvendas

autocorrelation plot

2.3 Requisitos

entender o negócio

3 Previsão

Now that we have a good idea of the problem in questian and all the data related to that problem, now we shall start to use Machine Lerning techniques to predict the values of the next day using all the data from the data in the excel file.

There are two way we can predict the values, using only one set of values, Univariate Search, or we can use more than one set of values, Multivariate Search. In this section of the document we will demonstrate the techniques used to predict and the respective results.

3.1 Univariados

For the Univariate Search we use two methods of predction, prediction using split and prediction using growing and rowling window.

3.1.1 Univariados Split

For the Split method first we split the number of occurences in two pieces, training set and test set, in our case it was a split of 80/20, that is 80% training and 20% test.

After that we used predictions techniques from the libraries rminer and forecast and we arrived at this conclusions:

3.1.2 Rowling and Growing Windows

For the Growin/Rowling Window methods consists in doing the split in windows, for our exemple we used windows of the size 7, that where meant to indicate 7 days, one week.

After that we used predictions techniques from the libraries rminer and forecast and we arrived at this conclusions:

3.2 Multivariados

3.2.1 Multivariados Split

4 Optimização