# Predictive Analytics with Factor Variance Association

Raul Ramirez-Velarde[1] and Laura Hervert-Escobar[1]and Neil Hernandez-Gress[1]

[1] Tecnológico de Monterrey NL Mexico
laura.hervert@tec.com

**Abstract.** Organizations are turning to predictive analytics to help solve difficult problems and uncover new opportunities. Nowadays, the processes are saturated in data, which must be used properly to generate the necessary key information in the decision making process. Although there are several useful techniques to process ans analyze data, the main value starts with the treatment of key factors. In this way, a Predictive Factor Variance Association (PFVA) is proposed to solve a multi-class classification problem. The methodology combines well-known machine learning techniques along with linear algebra and statistical models to provide the probability that a particular sample belongs to a class or not. It can also give predictions based on regression for quantitative dependent variables and carry-out clustering of samples. The main contribution of this research is its robustness to execute different processes simultaneously without fail as well as the accuracy of the results.

**Keywords:** PCA. Singular Value Decomposition, Machine Learning

## 1 Introduction

Machine learning has recently risen as one the most groundbreaking technologies of our times. Many companies are using machine learning and other data science techniques to improve processes and resource allocation improving significantly business value. While many Machine Learning algorithms have been around for a long time, the ability to automatically apply complex mathematical calculations to big data over and over, faster and faster is a recent development. Some of the challenges faced by companies is the high-dimensional data. Dealing with many variables can help in certain situations but also may divert attention from what really matters. Therefore, it is important to check whether the dimensionality can be reduced while preserving the essential properties of the full data matrix. Principal Component Analysis (PCA) is a most widely used tool in exploratory data analysis and in machine learning for predictive models. Moreover, PCA is an unsupervised statistical technique used to examine the interrelations among a set of variables. It is also known as a general factor analysis where regression determines a line of best fit. The main idea of this procedure is to reduce dimensionality of a dataset while preserving as much 'variability' (statistical information) as possible. Preserving variability may sometimes implies the finding of new variables that are linear functions of those in the original dataset, that successively maximize variance and that are uncorrelated with each other. Finding such new variables, the principal components (PCs), reduces to solving an eigenvalue/eigenvector

ICCS Camera Ready Version 2019
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-22750-0_28

problem. Mathematical basis of the methodology were presented by Pearson [1] and Hotelling [2]. The advances in technology for data processing have generated a broad study of PCA method. Literature is vast, some of the most substancial books for understanding PCA are [3,4], and for more specialized applications are [5,6]. Also, the use of machine learning techniques for prediction models are widely study in literature, Hervert [15] present a PCA method for reduction dimensionality combined with a multiple regression analysis to generate econometrics models. Then, such models are optimized to obtain the optimal price of a set of products. The model was tested in a case-study showing favorable results in profits for the pilot stores. Additionally, the literature provide articles that compile the advances and uses of prediction using machine learning prediction techniques. Sharma [7] present a survey of well-known efficient regression approach to predict the stock market price from stock market data based. Buskirk [16] provide a review of some commonly used concepts and terms associated with machine learning modeling and evaluation. The introduction also provides a description of the data set that was used as the common application example different machine learning methods. In this research we present a machine learning technique that given a matrix of explanatory variables value samples can produce predictions for quantitative dependent variables. It can also label the samples allocating one or several classes. This technique was tested to determine prices of articles for sale in convenience stores, to predict pollution contingencies, to determine leisure activities for tourists, to establish the probability of metastasis in cancer patients or the malignity of tumors in breast cancer patients, and other applications.

The procedure starts with a principal component analysis (PCA). This derives in several linear combinations of the original explanatory variables called principal components projections. These linear combinations are used to carry out a least-squares curve fitting to give predictions about variables of interest. Some applications implies a classification of the samples, mainly when the variables of interest have a definition of success or failure (within a thresh-old value). In such cases, point probabilities are computed for every value in the principal component projections. Then, the curve-fitting model will throw approximate probabilities for the value of a given combination of independent input variables.

The rest of the manuscript is organized as follows. The proposed methodology is presented in Section 2- The testing and analysis of the procedure are presented in section 3. Finally conclusions are given in section 4.

## 2 General procedure

The general outline of the technique follows. We are given a matrix $X$ with data in which rows are samples ($m$ samples) with numerical values and columns are variables (n variables). The explanatory variables are separated into matrix $E$ ($o$ variables) and

DOI: 10.1007/978-3-030-22750-0_28

the dependent variables (*p* variables), or variables of interest, are separated into matrix D. Furthermore *n*=*o*+*p*. Also, from PCA [8] samples scores and variables loadings, dimension reduction and clustering can also be carried out. Figure 1 shows the steps of the procedure.

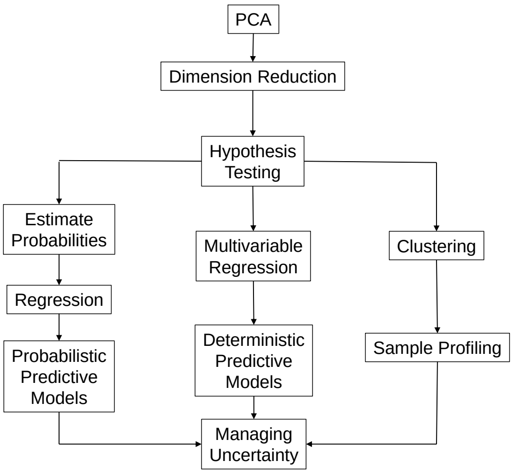

**Fig. 1.** Predictive Factor Variance Association Steps.

### 2.1 Dimension Reduction

The procedure starts by performing a PCA

1. Carry out PCA for matrix X and find matrices $X = PDQ^t$
2. Let $F = XQ$ be the principal component scores.
3. Use square cosines and a clustering algorithm over the first two columns of F (First two components) to determine:
   a. Collinearity and dimension reduction. Explanatory variables that are grouped together are collinear and thus eliminate those columns from matrix X. Repeat PCA, go to 1
   b. Causality relationships. If explanatory variables are grouped with variables of interest then that is evidence of a causality relationship.
   c. The square cosine is the square of the cosine of the angle between the vectors of variable loadings. If variables are close, the cosine will tend to one. If variables are separated, the cosine will approach zero. If $cosine^2(x) > 0.5$ then $cosine > \mp 0.7071$, thus associate variables with squared cosine greater than 0.5.
4. Eliminate variables of interest leaving only matrix E. Carry out PCA over E.
5. Select according to the following cases:
   a. **For predictive analytics** carryout curve fitting. Carry our curve fitting between explanatory variables and variables of interest clustered together in step 3.b. Discard any $R^2$<0.5. This indicates which explanatory variables influence the most the variable of interest

   b. **For multi-label problem** estimate probabilities. Match F$_i$ row by row with variable of interest j (assumed to be categorical with 0 or 1 as value), where i=1 … o and j=1…p. Use window procedure to estimate probabilities
6. Select according to the following cases:
   a. **For predictive analytics** match F$_i$ row by row with variable of interest j, where i=1 … o and j=1…p. Carry out curve fitting. Discard all R$^2$<0.5. Each F$_i$ is a linear combination of explanatory variables that potentially has enough information to give good predictions for variables of interest. Thus $\widehat{D_\iota} = f_{ij}(F_j) = f_{ij}(q_1X_1, q_2X_2, \dots, q_oX_o)$ where the coefficients q$_i$ are determined by matrix Q.
   b. **For multi-label problem c**arry out curve fitting. Discard all R$^2$<0.5. Each F$_i$ is a linear combination of explanatory variables that potentially has enough information to give good predictions for variables of interest. Thus $\widehat{D_\iota} = f_{ij}(F_j) = f_{ij}(q_1X_1, q_2X_2, \dots, q_oX_o)$ where the coefficients q$_i$ are determined by matrix Q. If $\widehat{D_\iota} > 0.5$ then assume a result of 1, and 0 otherwise.
7. Select according to the following cases:
   a. **For predictive analytics** normalized new samples called $X'$ can give predictions on variables, for $F' = X'Q$ and $\widehat{D_\iota'} = f_{ij}(F_j')$
   b. **For multi-label problem** normalized new samples called X’ can give predictions on variables, for $F' = X'Q$ and $\widehat{D_\iota'} = f_{ij}(F_j')$. Again, if $\widehat{D_\iota} > 0.5$ then assume a result of 1, and 0 otherwise.

### 2.2 Sorting and Estimating Probabilities

When a multi-label classification solution is required, class probabilities per sample must be calculated [9]. We introduce a new measure called the success probability which is usually interpreted as $P[x_i] > l$, that is the probability that the variable will reach a value above a threshold (although there are many other types of categories). The horizontal axis for this estimate can be any of the explanatory variables and the vertical axis any of the variables of interest which would in this case be categorical with a value of 1 for success or yes, and 0 for failure or no. But the most effective models are usually derived from using an entire factor, that is the linear combinations of explanatory variables obtained as columns of matrix $F$ to predict variables in matrix $D$.

To estimate this probability, for each factor and each variable of interest, an $m$-row vector of two entries is created. One entry is $vx$, or the explanatory linear combination from $F$ (or just single variable) and the other entry is $vy$, the categorical variable used to estimate probabilities. This vector is sorted by $vx$. A sample of $h$ items of is taken above and below a given value of $vx$ is taken creating a sliding window of samples. In general, $h$ data points from the plots are lost, $h/2$ points at the beginning of the plot an $h/2$ at the end. We use $h$=21 and $h = 41$. The samples are taken and then the number of successes are counted. So $vy_i$ is a measure which indicates that horizontal point $vx_i$is a success $(vy_i = 1)$ or not $(vy_i = 0)$. The probability of success $P_s(i)$ at horizontal point i is estimated by (Eq. 1):

ICCS Camera Ready Version 2019
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-22750-0_28

$$P_s(i) = \frac{1}{h} \sum_{j=i-\frac{(h-1)}{2}}^{j=i+\frac{(h-1)}{2}} vy_j \qquad (1)$$

In essence, we consider each success result in the sample as a Bernoulli experiment and use the sample mean of a collection of results of the variable of interest as an estimator for the probability, which is a well-known maximum likelihood estimator [10]. This is similar to Krichevsky-Trofimov [12] estimator, which is used as a conditional estimator of the outcome of the next Bernoulli experiment [11]. However, since our dataset is not a time series and our objective is to create a global model for the success probability as a function of some variable, we can use the sample mean of the current windows as seen in [12,13].

The sample size used to estimate the success probability is small, so tests were made with different values of $h$ such as 11, 41 and even 101 when possible, and it was found that the behaviour of the probability estimate was about the same, showing stability in the estimate. Naturally, the bigger the sample the more precise the estimate will be but the more variability will be lost to averaging.

# 3 Testing

In this section we present different data sets were the proposed methodology is tested.

## 3.1 Metastasis on Breast Cancer Patients

This data set consists of 2920 reference values for different genes of 78 breast cancer patients. These patients already have breast cancer tumors. We wish to determine the likelihood of tumors metastasizing. This is a binary classification problem. This data file was obtained from the UCI Machine Learning Repository (Dua and Karra 2017).

The probability estimate procedure with a windows of only 11 samples found a cubic equation with $R^2 = 0.831$. The fitted model with F17 is able to correctly determine if the sample metastasized or not on 75.64% of the samples. See figure 2 for a plot of probabilistic model.

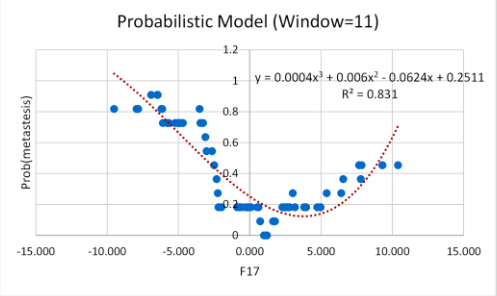

**Fig. 2.** Probabilistic model determining the probability of metastasis in breast cancer patients.

### 3.2 Malignity of Tumors in Breast Cancer Patients

This data set consists of 570 samples of information about cancer patients. The variables are information about the tumor such as radius, perimeter, area, texture, softness, concave points, concavity, symmetry and fractal dimension. The objective is to find if the tumor is malignant or benign. This is a binary classification problem. This data file was obtained from the UCI Machine Learning Repository (Dua and Karra 2017).

The fitted probabilistic model using F1 shows a coefficient determination of $R^2 = 0.9434$ and is able to determine correctly if the sample corresponds to a malignant tumor in 91.38% of the samples. See figure 3 for a plot of the probabilistic model.

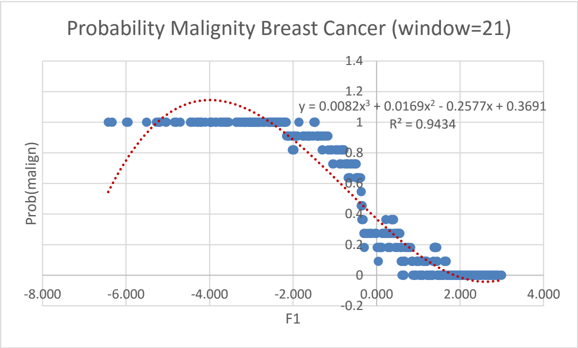

**Fig. 3.** Probabilistic model determining if a tumor is malignant in breast cancer patients.

**Dimension Reduction.** PCA on the data set also shows that some of the variables can be eliminated achieving dimension reduction. This is shown in figure 4, where some variables can be eliminated. This is very important. If a medical device is being designed to help people heal or prevent a health problem, the less variables that are necessary to control the cheaper the device will be and the more people it will help.

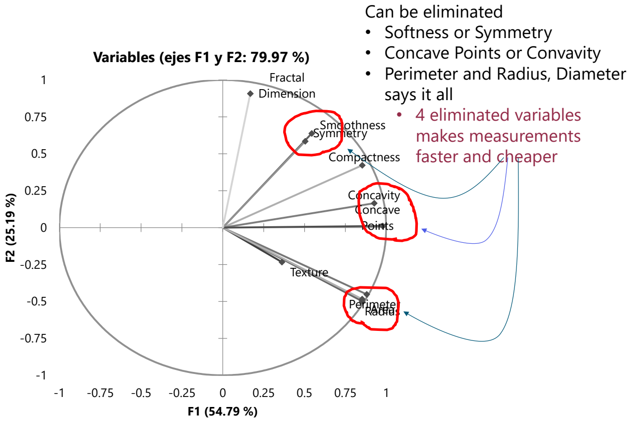

**Fig. 4.** Dimension reduction to determine of a tumor is malignant

DOI: 10.1007/978-3-030-22750-0_28

Figure 5 shows how the clustering in the F1 vs F2 biplot of samples agrees with using F1 to separate malignant to benign tumors. As can be seen, there is a clear separation between samples marked as M or B along the F1 axis.

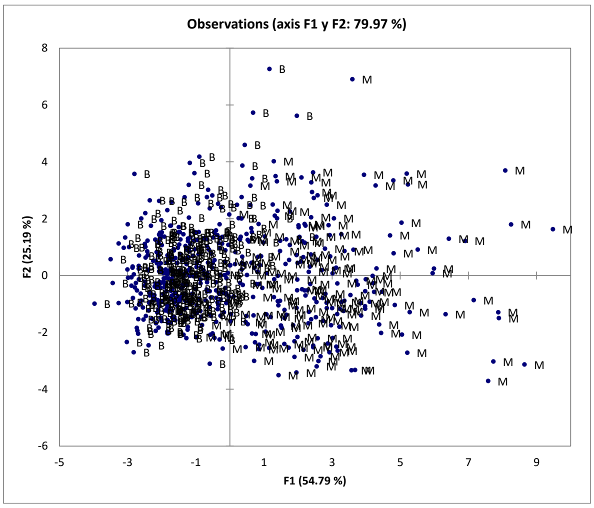

**Fig. 5.** Clustering of samples. M means malign, B means benign.

### 3.3 Travel and Activities

This data set corresponds to 87 samples containing answers to a survey. The provided information is the individual psych-socio-economic profile. This data is useful for traveling websites in order to determine what kind of activities a user would prefer. The information contains several features for the socio-economic profile, and for psychological profile.

The socio-economic features are gender, age, education, marital, employment, WEMWBS, PANAS: PA, PANAS: NA, SWLS, SWLS: group.

The features for psychological profile are:

- Personality traits: Extroversion, Agreeableness, Consciousness, Imagination, Neuroticism shown as the columns: 'Big5: Extraversion', 'Big5: Agreeableness', 'Big5: Conscientiousness', 'Big5: Neuroticism', 'Big5: Imagination'
- 3 orientations to happiness: Pleasure, Meaning, and Engagement. The columns: 'OTH: Pleasure', 'OTH: Meaning', 'OTH: Engagement'
- Fear of missing out (FoMO).

The types of activities are (Would the person like to do…?): 'Outdoors-n-Adventures', 'Tech', 'Family', 'Health-n-Wellness', 'Sports-n-Fitness', 'Learning', 'Photography', 'Food-n-Drink', 'Writing', 'Language-n- Culture', 'Music', 'Movements', 'LGBTQ', 'Film' , 'Sci-Fi-n-Games', 'Beliefs', 'Arts', 'Book Clubs', 'Dance', 'Hobbies-n-

Crafts', 'Fashion-n-Beauty', 'Social', 'Career-n-Business', 'Gardening-n-Outdoor housework', 'Cooking', 'Theatre, Show, Performance, Concerts', 'Drinking alcohol, Partying', 'Sex and Making Love'.

The data file was obtained from online surveys.

This is a multi-label problem. In table 1 we show which activities are predicted by which factors with and the model $R^2$.

**Table 1.** Types of activities by component best able to predict it and model R$^2$.

| Variable | F | R2 | Variable | F | R2 | Variable | F | R2 |
|---|---|---|---|---|---|---|---|---|
| **1** | 16 | 0.748 | 11 | 7 | 0.725 | 21 | 3 | 0.680 |
| **2** | 19 | 0.609 | 12 | 1 | 0.774 | 22 | 2 | 0.710 |
| **3** | 9 | 0.674 | 13 | 10 | 0.651 | 23 | 11 | 0.704 |
| **4** | 14 | 0.769 | 14 | 13 | 0.660 | 24 | 15 | 0.676 |
| **5** | 15 | 0.670 | 15 | 17 | 0.728 | 25 | 19 | 0.591 |
| **6** | 8 | 0.711 | 16 | 4 | 0.812 | 26 | 13 | 0.647 |
| **7** | 7 | 0.727 | 17 | 16 | 0.729 | 27 | 19 | 0.676 |
| **8** | 16 | 0.715 | 18 | 7 | 0.766 | 28 | 13 | 0.562 |
| **9** | 20 | 0.690 | 19 | 15 | 0.642 | | | |
| **10** | 4 | 0.945 | 20 | 7 | 0.693 | | | |

For example Activity 'Language-n-Culture' can be predicted by the linear combination of explanatory variables formed by factor 4, with and $R^2 = 0.945$. The model is able to determine accurately that a person will like this type of activity in 67.81% of the samples. Figure 6 shows the probabilistic model.

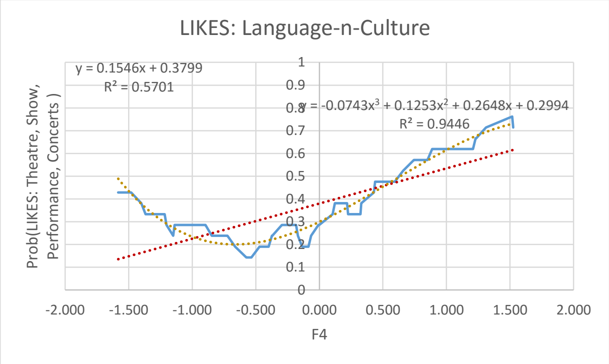

**Fig. 6.** Probabilistic model for language and culture related activities preferences.

### 3.4 Convenience Store Pricing Strategy

This is a data set of 6362 samples of sales tickets on different stores of several convenience stores. The data consists of type of merchandise, merchandize identification number, year and year-week, article cost, taxes, units bought, purchase amount, units with discount, amount of discount, articles prize, total profit margin and % of profit margin (from cost), minimum, medium and maximum weather temperatures, and amount of rain.

Principal component analysis determined that profit margin and % of profit margin as well as units sold is dependent only in the units sold with discount and the total amount of discount. See figure 7.

For the convenience store, one important goal is to sale articles with profit margin of at least 40%. PFVA determined that F10 of a PCA that excludes variables "profit margin" and "% of profit margin" gave the best prediction about attaining this goal with an $R^2 = 0.753$. The regression model is able to determine if the sales ticket will achieve the goal in 68.86% of the samples.

An important observation is that profit margin and the probability of achieving more than 40% of profit margin are opposite. This is because margin increases with the number of articles sold at discount but the probability of achieving more than 40% profit margin diminishes, as discount articles have lower price at same cost than regular priced articles. That can be seen in figures 8a and b for profit margin and Prob(%margin>40%) versus discount units sold (usdes) and figures 9a and b for profit margin and Prob(%margin>40%) versus discount total amount (montodes), both for article identifies as sku=455.

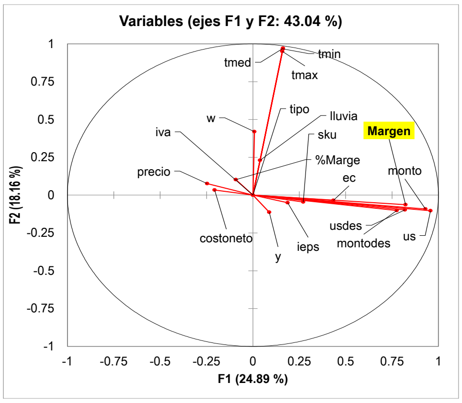

**Fig. 7.** Principal Component Analysis of convenience store sales

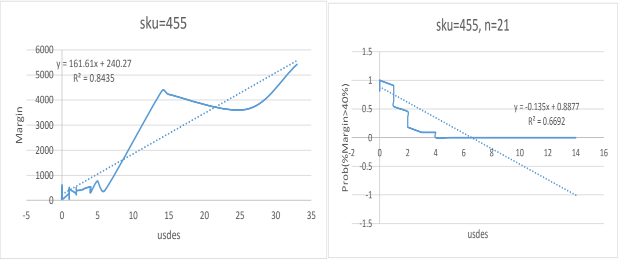

**Fig. 8.** a. Profit margin versus discount units sold for article sku=455. b. Prob(%profit margin>40%) versus discount units sold for article sku=455

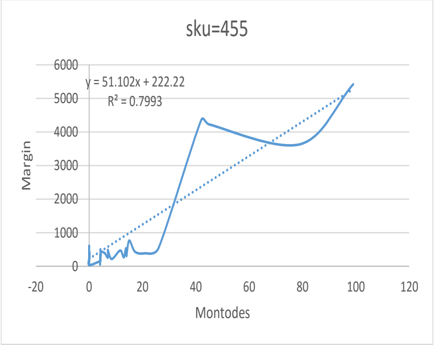

**Fig. 9.** a. Profit margin versus discount units sold for article sku=455

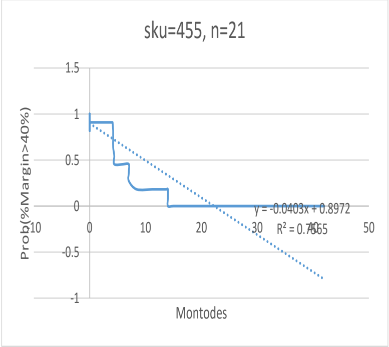

**Fig. 9.** b. Prob(%profit margin>40%) versus discount units sold for article sku=455

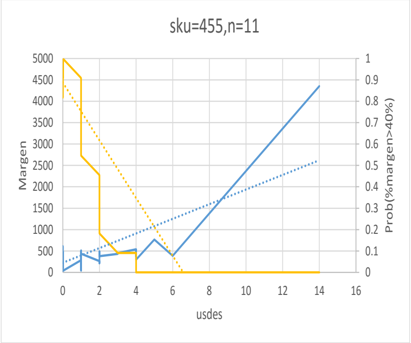

**Fig. 10.** a. Balance between Profit margin versus total margin. Allow at most 5 units sold at discount per sale

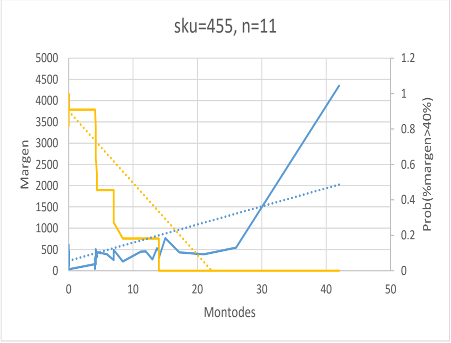

**Fig. 10.** b. Balance between Profit margin versus total margin. Allow at most 16 currency units discount per sale

### 3.5 Metropolitan Area Air-Pollution

Monterrey's Metropolitan Area (MMA) in Mexico, consists of 15 municipalities with an approximate population of 4,406,054 inhabitants. As it so happens with other large metropolitan areas, pollution is a concern. More than 60% of days in a year have pollution levels that label air quality as bad or extremely bad. There are 12 monitoring stations throughout the metropolitan area, although for this paper we only used data from five stations because the other had too much incomplete data.

Each monitoring station measures every hour weather variables such as pressure (PRS), temperature (TOUT), relative humidity (HR), solar radiation (SR), rainfall (Rain), wind speed (WSR) and direction (WDV); and pollutants such as carbon oxide (COx), nitrogen oxide (NOx), sulfur oxide (SOx), ozone (O3), particles with diameter less than 2.5 microns (PM2.5) and particles with diameter less than 10 microns (PM10). Regional health authorities consider the last three pollutants the ones that have the most adverse effects on population. Nevertheless, in only 3% of 2015 measured days did daily O3 maximum concentration exceed norms, whereas PM10 and PM2.5 exceed maximum limits in 58% and 63% of days respectively. Although PM10 includes PM2.5 particles, usually PM10 particles greater than 2.5 microns (but less than 10 microns) are mainly dust. There are in total 14,374 samples. The information was obtained from Integral Air Quality Monitoring system from Nuevo Leon province in Mexico (Martinez et al, 2012).

**PM2.5 (Particles with less than 2.5 microns in diameter).** Particulate matter, or PM, is the term for particles found in the air. Many manmade sources emit PM directly or emit other pollutants that react in the atmosphere to form PM (Martinez et al, 2012). PM10 pose a health concern because they can be inhaled into and accumulate in the respiratory system. PM2.5 are referred to as "fine" particles and are believed to pose the greatest health risks. Because of their small size fine particles can lodge in the respiratory system, and may even reach the bloodstream. Exposure to such particles can affect respiratory and cardiovascular systems. Numerous scientific studies have linked particle pollution exposure to a variety of problems, including (Lerma et al, 2013): premature death in people with heart or lung disease, nonfatal heart attacks, irregular heartbeat, aggravated asthma, decreased lung function, increased respiratory symptoms, such as irritation of the airways, coughing or difficulty breathing.

Sources of fine particles include all types of combustion activities and industrial processes but also they are indirectly formed when gases from burning fuels react with sunlight and water vapour (Martinez at al, 2016). These can result from fuel combustion in motor vehicles, at power plants, and in other industrial processes. Most particles form in the atmosphere as a result of complex reactions of chemicals such as sulphur dioxide and nitrogen oxides, which are pollutants emitted from power plants, industries and automobiles.

PM2.5 is the main concern since these particles form out of dangerous chemicals that have very serious effects on population. Several measures have been proposed to reduce PM2.5 pollution such as restricting vehicle transit by license plate number (that is, some vehicles would not be allow to circulate some days of the week) and vehicle verification. As we can see in figures 11a and b, although ozone can be directly related to vehicle traffic (as represented by time of day) PM2.5 cannot (Carrera et al, 2015). Therefore, ***none of those vehicle related preventive measures would work***.

DOI: 10.1007/978-3-030-22750-0_28

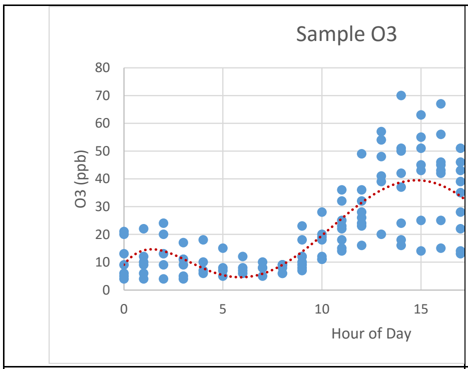

**Fig. 11.** a. Changes in O3 air concentration according to time of day. O3 is clearly related to daily traffic as represented by time of day

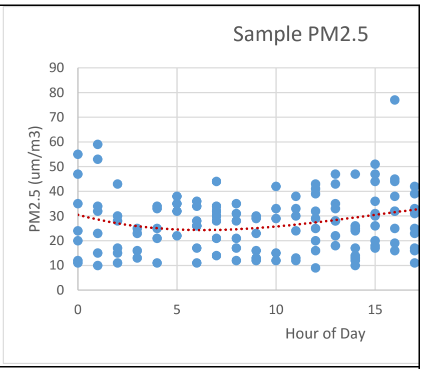

**Fig. 11.** b. Changes in PM2.5 air concentration according to time of day. PM2.5 is clearly NOT related to daily traffic as represented by time of day

Also, weather affects pollution. Our aim is to, given a morning weather forecast, to determine if there will be a pollution contingency, that is, pollution levels higher than allowed by laws and regulations. This would allow regional governments to implement emergency measures to protect the population.

PCA shows (See figure 12) that O3 is mainly related to temperature, solar radiation and relative humidity, as it's well known that O3 levels are often high when the day is hot, sunny and dry. Nevertheless O3 levels can also be high at low temperatures.

For O3 we are able to create a deterministic model as shown in figure 13. F1 can predict the value of O3 with $R^2 = 0.7801$.

Figure 12 also shows that PM2.5 and PM10 are very weakly related to weather, as the biplot lines are orthogonal from any weather variable. Nevertheless, since we are mainly worried about the daily maximum levels, a new file was created with 365 samples containing daily maximum levels of PM2.5 and maximum, minimum and average daily readings for all weather variables. We found F10 from this new analysis to be best predictor of the class PM2.5>40.5. It has $R^2 = 0.9663$ and it's able to correctly predict bad air quality because of high PM2.5 levels in 65.7% of the samples, giving evidence that even though weather does influence pollution, for air suspended particulate neither weather nor traffic are the main determining factors. Since the number of samples was large, we tried windows from 11 samples to 101 samples, finding that the windows from 11 to 51 would have the same percentage of correct predictions, with that percentage dropping slightly at window size 101.

DOI: 10.1007/978-3-030-22750-0_28

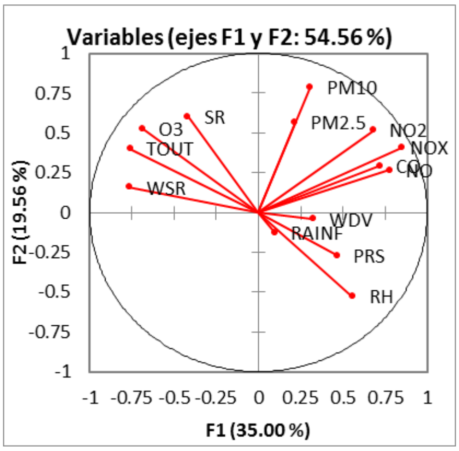

**Fig. 12.** PCA indicating that O3 is related to weather, specifically temperature, humidity and solar radiation. There seem to be no clear relationship between weather and particles.

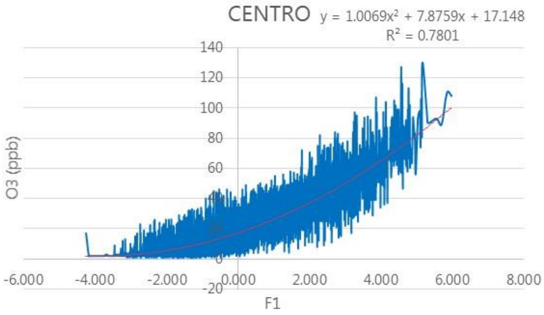

**Fig. 13.** Predictive model for O3 based on F1.

## 4 Conclusions

PFVA is a new technique that uses well-known concepts of matrix algebra and probability that is able to solve several problems of data analytics. It provides orthogonal linear combinations of the explanatory variables that can be used to predict the value of a variable of interest given a collection of values for explanatory variables; determine the classes to which a sample belong; and to classify samples into groups with common characteristics.

The main differentiator of this technique is the use of sample window to compute class probabilities that is has proven to be, even though the window can be really small, as seen in the examples presented, to be robust and accurate.

ICCS Camera Ready Version 2019
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-22750-0_28

## References

1. Pearson K.. On lines and planes of closest fit to systems of points in space. Phil. Mag. 2, 559–572. 1901 (10.1080/14786440109462720)
2. Hotelling H.Analysis of a complex of statistical variables into principal components. J. Educ. Psychol. 24, 417–441, 498–520 1993 (10.1037/h0071325)
3. Jackson JE. A user's guide to principal components. 1991 New York, NY: Wiley.
4. Jolliffe IT.. Principal component analysis, 2nd edn New York, NY: Springer-Verlag. 2002
5. Diamantaras KI, Kung SY. Principal component neural networks: theory and applications. New York, NY: Wiley. 1996.
6. Flury B. Common principal components and related models. New York, NY: Wiley. 1988
7. A. Sharma, D. Bhuriya and U. Singh, "Survey of stock market prediction using machine learning approach," 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, 2017, pp. 506-509.
8. Jolliffe, I. T. .Principal Component Analysis and Factor Analysis. In: Principal Component Analysis. Springer Series in Statistics. Springer, New York, NY 2012
9. Golub, Gene H., and Charles F. Van Loan. Matrix computations. Vol. 3. JHU Press, 2012.
10. Wilks, S.S. (1962), Mathematical Statistics, New York: John Wiley & Sons. ISBN 978-0471946502.
11. Belyaev, E., Gilmutdinov, M., & Turlikov, A. (2006). Binary Arithmetic Coding System with Adaptive Probability Estimation by "Virtual Sliding Window." 2006 IEEE International Symposium on Consumer Electronics, 1–5. https://doi.org/10.1109/ISCE.2006.1689517
12. Krichevsky, R., & Trofimov, V. (1981). The performance of universal encoding. IEEE Transactions on Information Theory, 27(2), 199-207. 10.1109/TIT.1981.1056331
13. Leighton, F., & Rivest, R. (1986). Estimating a probability using finite memory. IEEE Transactions on Information Theory, 32(6), 733-742. 10.1109/TIT.1986.1057250
14. Dua, D. and Karra Taniskidou, E. (2017). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.
15. Laura Hervert-Escobar, Oscar A. Esquivel-Flores, Raul V. Ramirez-Velarde, Optimal pricing model based on reduction dimension: A case of study for convenience stores,Procedia Computer Science,Volume 108, 2017
16. Buskirk, Trent & Kirchner, Antje & Eck, Adam & S. Signorino, Curtis. (2018). An Introduction to Machine Learning Methods for Survey Researchers. Survey Practice. 11. 1-10. 10.29115/SP-2018-0004.

ICCS Camera Ready Version 2019
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-22750-0_28