ISBN: 978-628-95207-4-3. ISSN: 2414-6390. Digital Object Identifier: 10.18687/LACCEI2023.1.1.832

Utilización de un marco de inteligencia de negocios y minería de datos como herramientas computacionales en las PYMES: Pronóstico de producción de una planta de energía hidroeléctrica como caso de estudio

Redy H. Risco-Ramos, Eng.D.[1], Daniel A. Pérez-Aguilar, Eng.D.[2,3], Luis A. Casaverde-Pacherrez, Eng.D.[1], Manuel Malpica Rodríguez, Dr.[2], Jair M. Pérez-Aguilar, M.Sc.[2,3], and Andy P. Pérez-Aguilar, M.Sc.[3]
[1]Universidad de Piura - Piura, Perú
[2]Universidad Privada del Norte (UPN) - Cajamarca, Perú
[3]Universidad Tecnológica del Perú (UTP) - Lima, Perú
redy.risco@udep.edu.pe, daniel.perez@upn.pe, luis.casaverde@alum.udep.edu.pe, manuel.malpica@upn.pe, n00030252@upn.pe, c25932@utp.edu.pe

***Resumen*– El objetivo de este artículo es demostrar la utilidad de utilizar el modelo de referencia CRISP-DM (Cross-Industry Standard Process for Data Mining) como metodología de minería de datos dentro de una plataforma comercial de inteligencia de negocios (BI) para pequeñas y medianas empresas (PYMES). Para lograr este objetivo se aplicó el modelo CRISP-DM, se utilizó como plataforma de BI la aplicación Power BI y se implementó la metodología en un caso de estudio de uso específico dentro de una PYME de generación hidráulica. Los resultados muestran que la metodología se puede aplicar con hardware y software fácilmente disponibles, lo que genera mejores resultados en comparación con las prácticas convencionales utilizadas en la empresa. Con base en estos resultados, se puede concluir que el método utilizado es de suma utilidad, factible de implementar con hardware y software accesibles para las PYMES, y el modelo de referencia sigue siendo válido cuando los objetivos de minería de datos están bien definidos.**

***Palabras clave*— Inteligencia de negocios, CRISP-DM, Data mining, Pronóstico.**

Digital Object Identifier: (only for full papers, inserted by LACCEI).
ISSN, ISBN: (to be inserted by LACCEI).
DO NOT REMOVE

Using a business intelligence framework and data mining as computational tools in SMEs: Production forecasting of a hydroelectric power plant as a case study

Redy H. Risco-Ramos, Eng.D.[1], Daniel A. Pérez-Aguilar, Eng.D.[2,3], Luis A. Casaverde-Pacherrez, Eng.D.[1], Manuel Malpica Rodríguez, Dr.[2], Jair M. Pérez-Aguilar, M.Sc.[2,3], and Andy P. Pérez-Aguilar, M.Sc.[3]
[1]Universidad de Piura - Piura, Perú
[2]Universidad Privada del Norte (UPN) - Cajamarca, Perú
[3]Universidad Tecnológica del Perú (UTP) - Lima, Perú
redy.risco@udep.edu.pe, daniel.perez@upn.pe, luis.casaverde@alum.udep.edu.pe, manuel.malpica@upn.pe, n00030252@upn.pe, c25932@utp.edu.pe

Abstract*–The aim of this paper is to demonstrate the utility of using the CRISP-DM (Cross-Industry Standard Process for Data Mining) reference model as a data mining methodology within a commercial business intelligence (BI) platform for Small and Medium Enterprises (SMEs). To achieve this objective, the CRISP-DM model was applied, the Power BI application was used as a BI platform, and the methodology was implemented in a specific use case study within a hydraulic generation SME. The results show that the methodology can be easily applied with readily available hardware and software, resulting in better outcomes compared to conventional practices used in the company. Based on these results, it can be concluded that the method used delivers useful results, it is feasible to implement with accessible hardware and software for SMEs, and the reference model remains valid when the data mining objectives are well-defined.

Keywords--Business Intelligence, CRISP-DM, Data Mining, Forecasting.

Resumen*– El objetivo de este artículo es demostrar la utilidad de utilizar el modelo de referencia CRISP-DM (Cross-Industry Standard Process for Data Mining) como metodología de minería de datos dentro de una plataforma comercial de inteligencia de negocios (BI) para pequeñas y medianas empresas (PYMES). Para lograr este objetivo se aplicó el modelo CRISP-DM, se utilizó como plataforma de BI la aplicación Power BI y se implementó la metodología en un caso de estudio de uso específico dentro de una PYME de generación hidráulica. Los resultados muestran que la metodología se puede aplicar con hardware y software fácilmente disponibles, lo que genera mejores resultados en comparación con las prácticas convencionales utilizadas en la empresa. Con base en estos resultados, se puede concluir que el método utilizado es de suma utilidad, factible de implementar con hardware y software accesibles para las PYMES, y el modelo de referencia sigue siendo válido cuando los objetivos de minería de datos están bien definidos.

***Palabras clave*—Inteligencia de negocios, CRISP-DM, Data mining, Pronóstico.**

Digital Object Identifier: (only for full papers, inserted by LACCEI).
ISSN, ISBN: (to be inserted by LACCEI).
DO NOT REMOVE

I. INTRODUCTION

At present, small and medium-sized enterprises (SMEs) [1] in Peru still do not take advantage of the great value that the enormous amount of data available as a result of their operations can provide. This is so despite the fact that, as stated in [2] , since 1865 there have been processes, architectures and technologies that can transform this data into relevant information and knowledge for making important decisions and therefore for the efficient management of their businesses.

This set of processes, architectures and technologies is known as business intelligence (BI) [3][4]. Another concept closely related to BI is Data Mining (DM) which is a subdomain of Artificial Intelligence defined as a process that aims to generate knowledge from data presenting the findings in an integral way to the user [5][6].

The generation of knowledge in the context of DM can translate into the discovery of patterns, relationships, and trends relevant to the user. DM as a process involves, in essence, the collection and selection of data, the preprocessing of data, the analysis of data itself, including the visualization of results, the interpretation of findings, and the application of the knowledge obtained. To pre-process and analyze the data, machine learning (ML) and statistical methods are implemented [5][7][8].

CRISP-DM (Cross Industry Standard Process for Data Mining) is a widely used framework for planning and executing data mining projects. This process is structured into six phases: Business and Data Understanding, Data Preparation, Data Analysis, Modeling, Evaluation, and Deployment. CRISP-DM provides a step-by-step guide for conducting data mining projects efficiently and organizedly [9], [10].

In [11], a testing method for BI projects based on software engineering tests is proposed. It is important to include verification and validation practices to test the quality and success of the project. Additionally, it is necessary to establish permanent tools and mechanisms to analyze the completeness and consistency of the BI solution construction. Including the process assets and best practices identified in project management is also important. The proposed model includes a post-mortem analysis at the end of each stage to identify key elements and, if necessary, return to a previous stage to make adjustments to the solution.

In [12], the implementation of a business intelligence system in agricultural production is studied, where producers base their decisions on unofficial information. The research was quasi-experimental and registered producers in a public institution. The results show that the implementation of the system reduced the response time to queries and improved the reach of information, from 7% to 71%. Most producers are in agreement with the implementation of the system.

The objective of [13] was to determine the acceptance of automotive products in the market. A database of 1728 records is used and the influence of 6 independent variables on acceptance is analyzed using the J48 data mining algorithm on the Weka platform. The results show that the algorithm is capable of predicting the acceptance behavior with an effectiveness of 92% and that the most influential variables are cost, number of people, and safety.

Another study was carried out with the objective of improving demand forecasting models for a convenience store chain. The company detected a significant increase in sales during a heat wave in 2018, so it decided to improve the forecasting models and evaluate the importance of external variables. Different forecasting models were created and validated, and finally, two were selected to propose a forecasting method and a final application that improves the company's logistics planning and decision-making [14].

In [15], an extension to the CRISP-DM process model that supports data mining applications in engineering projects is presented. The extension, called DMME, was developed based on the experiences of numerous previous projects in the machine tool and adaptive control development chair. It is indicated that DMME provides support in data acquisition and provides context information about the data, while maintaining the advantages of CRISP-DM in the data mining part. It is concluded that it is a structured and focused approach for data mining projects in the field of engineering, allowing engineers, data analysts, and computer scientists to work together to achieve the goals.

In [16], a process model for the development of machine learning applications is presented, covering six phases from scope definition to the maintenance of the implemented application. Special attention is paid to the last phase, as a model in constantly changing real-time environments requires close monitoring and maintenance to reduce the risk of performance deterioration over time. For each task in the process, an appropriate quality assurance methodology is proposed to tackle the challenges identified as risks in the development of machine learning, extending the CRISP-DM model.

In [17], it is noted that Small and Medium Enterprises (SMEs) need to adopt data analytics. Despite data mining (DM) being widely used in the transportation sector, there are surprisingly few case studies on DM research. The three most common DM models used by large companies in the transportation sector, including CRISP-DM, are identified. Concluding that there is a critical need to develop a novel model to meet the requirements of SMEs.

In Peru, the use of data mining algorithms, business intelligence, prediction models, etc. In SMEs is limited by the lack of specialists in these companies and the inaccessibility of some of these tools.

This work proposes to use a highly validated methodology, CRISP-DM, in a BI tool accessible to any SME for commercial use. The aim is to demonstrate how prediction can be improved in a small hydroelectric generation company by using known prediction models that are transparent to the user, based on simple averages of past productions.

Keep in mind that it is not the objective of this work to optimize the prediction, since obviously it would be necessary to adjust parameters of the neural networks, test other prediction models and this is not the scope of this work.

Currently, data analytics, BI and DM platform providers are increasing including capabilities within their platforms that go far beyond their traditional markets. Analytics and BI platforms increasingly include functionality to perform augmented DM tasks. Predictive models run without having direct contact with the user, but delivering information and insights appear naturally within the BI analysis and process flow [18].

Next, we describe the development of this work: in section 2 we present the methodology, in section 3 we present the discussion of the results, and finally, in section 4 we present some conclusions of the work.

II. Methodology

The objective of this work is achieved through the following methodology: Firstly, the CRISP-DM reference model is selected and presented as the tool for an SME to develop a data mining solution. Then, the commercial business intelligence platform (Power BI) is briefly introduced, which will be used in accordance with the chosen reference model. Finally, the proposal is tested in a case study.

1) The CRISP-DM (Cross-Industry Standard Process for Data Mining) model.

For the development of a Data Mining solution in 1996, four leaders of the nascent DM market conceived the CRISP-DM (Cross-Industry Standard Process for Data Mining) model, as a non-proprietary, documented, and freely available model that allowed organizations to achieve better data mining results and promote best practices in the industry. This model is described in [19][20].

Fig. 1 shows the phases of a DM process, proposed by CRISP-DM, in which the six phases are as follows: "understanding the business, understanding the data, data preparation, modeling, evaluation, and deployment" [21].

In [20] has provided an explanation for Fig. 1, in which the sequence of steps and the interrelation between them can be observed. Additionally, the cyclical nature of the process is represented by the large ellipse. This indicates that after some solutions have been identified based on the knowledge found, new solutions can be generated from the previous discoveries.

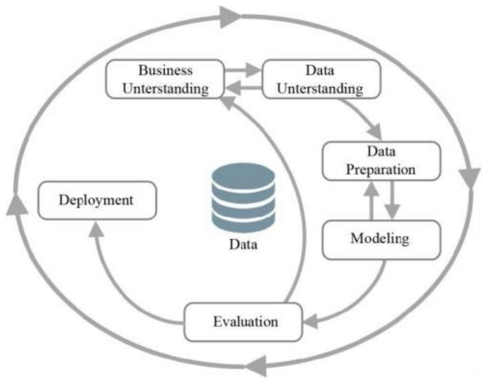

Fig. 1. Phases of the CRISP-DM reference model [21].

Martinez-Plumed et al. in [22] has done a very good analysis of CRISP-DM after 20 years of use and states that CRISP-DM is still considered the most comprehensive data mining methodology in terms of meeting the needs of industrial projects, and it has become the most widely used process for DM projects. In short, CRISP-DM is considered the de facto standard for analytics, data mining, and data science projects.

However, in [22] also discusses new alternatives to the CRISP-DM model, taking into account the expanding scope of data science. Nevertheless, the conclusion is that when a process has a well-defined business objective and the necessary data has already been collected and is available for further computational processing, CRISP-DM remains a suitable methodology to follow and still holds up after two decades.

For these reasons we propose to use this reference model as a methodology to implement a Business Intelligence framework applied to electricity companies or electricity users, especially small or medium-sized companies and that we describe in the following sections. In [9] the model is presented and summarizes its phases very well. Table II presents a summary of the phases of the reference model according to [9].

Testing the validated methodology in a small and medium-sized enterprise (SME) is valid because an SME lacks the personnel, technological resources, and digital capabilities that a large company has.

2) The business intelligence framework.

For the implementation of the data mining process, we chose a commercial tool that is easily accessible to an SME, such as Microsoft Power BI.

The Gartner Magic Quadrant for March 2022 identifies Microsoft Power BI as the leading analytics and business intelligence platform. One of its major strengths is its seamless integration and compatibility with Office 365, a widely adopted office software solution among Small and Medium Enterprises (SMEs).

Power BI is already being used in different case studies such as in [23], [24] o [25].

This tool has the advantage of being very accessible to SMEs, as it has a free use option and if more benefits are desired, a low-cost license would allow these benefits to be used. For example, the analyzes in this work have been carried out with a free Power BI license.

TABLE I
CRISP-DM PHASES

Phase	Description
Understanding the business	Evaluation and/or identification of the situation of the business or part of the business to obtain an overview of the available and necessary resources. Determination of the objective of data mining by defining the type of data mining to be applied and the success criteria of data mining. Creation of a project plan.
Understanding the data	Data collection from its sources, exploration, description, and verification of data quality. Statistical analysis is used to determine attributes and their collations
Data preparation	Selection of data by defining inclusion and exclusion criteria. Poor data quality can be handled by cleaning the data. Depending on the model used (defined in the first phase) the derived attributes must be constructed. For all these steps different methods are possible depending on the model
Modeling	It consists of selecting the modeling technique, and building the test case and the model. All data mining techniques can be used. In general, the choice depends on the business problem and the data. The important thing is how to explain the choice. To build the model, specific parameters must be set. To evaluate the model, it is appropriate to evaluate the model against the evaluation criteria and select the best
Evaluation	The results are compared with the defined business objectives. Therefore, the results must be interpreted and according to the interpretation, additional actions must be defined. This might require reviewing the overall process.
Implementation	It can be a final report or a software component. The user guide describes that the deployment phase consists of planning deployment, monitoring, and maintenance.

3) The case study.

Peru, since approximately 2010, electric generators with renewable resources, including hydroelectric plants with installed capacities of up to 20 MW, have been entering commercial operation in the Peruvian electricity system [13].

These plants have the requirement to report the estimated production of the following year to the regulatory, operator, and regulatory entities.

In addition, this estimated production is very useful for the projection of economic income for the following year. Generating companies usually make an estimate based on the averages of previous years.

For the application of the methodology to this specific case study we have followed the following steps:

The raw data was obtained from the production files of the generating company for one of its plants, the data was reviewed and preprocessed to give it an adequate model. In this case, the methodology is being applied to quite specific objectives, but the idea is to validate that the steps can be followed, and this level of useful knowledge can be obtained for decision making.

The methodology was applied with the available data of the company, using Power BI, as in [23], for the phases of data modeling, extraction, transformation, and loading of the data. For the transformations, the M language was used (transparent when implemented from the application interface) and the new measures or columns were generated in DAX language.

In Power BI, the prediction models embedded in the platform were implemented, the visualizations were also developed in Power BI and an evaluation of the results was made. The characteristics of the hardware used for the development of this work are shown in Table II. We proceed to apply the methodology.

TABLE II
HARDWARE USED

Processor	Intel(R) Core(TM) i7-9750HF CPU @ 2.60GHz 2.59GHz.
Installed RAM	16.0 GB
System Type	64-bit operating system, x64-based processor
Operating System	Windows 10 Home Single Language

Step One: Understanding the Business:

The company is a concessionary electricity generation company that must project the production of its generation plants annually. We arbitrarily chose one of their plants with 12.5 MW of installed capacity. The plant has been operating since 1998. The final objective is to estimate the monthly production of the following year.

The types of analysis to be carried out would be a descriptive one to visualize the past behavior and a predictive one to be able to estimate the production of the following year.

PRODUCCIÓN HISTÓRICA MENSUAL DE LA PLANTA A : DICIEMBRE 2021

		ENERO	FEBRERO	...	NOVIEMBRE	DICIEMBRE	ACUMULADO TOTAL MWh	ACUMULADO TOTAL GWh
AÑO 1998	MWh	790	1785	...	4484	5160	49562	50
AÑO 1999	MWh	2290	5418	...	1023	1516	60806	61
AÑO 2000	MWh	5288	5122	...	3598	3401	64033	64
AÑO 2001	MWh	5867	7165	...	3287	3757	77903	78
AÑO 2002	MWh	4048	4713	...	4441	3720	63763	64
AÑO 2003	MWh	4730	4328	...	2734	2776	48306	48
AÑO 2004	MWh	5734	4043	...	1021	2221	38234	38
AÑO 2005	MWh	5372	4772	...	2599	1965	48643	49
AÑO 2006	MWh	2441	7437	...	165	4414	44982	45
AÑO 2007	MWh	5732	7460	...	2073	3403	48843	49
AÑO 2008	MWh	4843	7539	...	4466	3131	60524	61
AÑO 2009	MWh	7033	7574	...	3559	2868	71523	72
AÑO 2010	MWh	4235	7071	...	2581	1963	54347	54
AÑO 2011	MWh	2628	5797	...	2169	5195	52096	52
AÑO 2012	MWh	8541	7618	...	3266	2650	65871	66
AÑO 2013	MWh	6718	7237	...	2013	1993	42535	43
AÑO 2014	MWh	1409	1935	...	2760	2341	41794	42
AÑO 2015	MWh	2184	5750	...	2437	2878	55761	56
AÑO 2016	MWh	1266	6245	...	1480	1296	47457	47
AÑO 2017	MWh	1422	7099	...	2354	1671	55620	56
AÑO 2018	MWh	2771	5906	...	767	279	39136	39
AÑO 2019	MWh	2703	6018	...	2856	1938	46815	47
AÑO 2020	MWh	2114	6223	...	1935	2483	46853	47
Promedio		3920	5837	...	2525	2740	53279	53
AÑO 2021(Real)	MWh	2508	6625	...	2910	1221	43708	44

Fig. 2. Original data. Monthly historical production.

Step two: Understanding the data.

Data comes from a file generated by operations management and updated month by month. Fig. 2 shows the month-to-month production of the analyzed plant. The format is the source format.

Step three: Prepare the data.

The data required format transformation to adapt it to a time series which can be seen in Fig. 3. In addition, an additional power plant table was created to generate a single vertex star model and a measurement table. The model related two tables by plant code and a measurement table was created. See Fig. 4.

CODIGO_CENTRAL	MES_YEAR	PRODUCCIÓN_MWh	Potencia Instalada	MES	YEAR	Inicio Mes	Fin de Mes	Num Dias	KPI_Factor de Planta	Mes Corto	FP_Proyectado
CHC	enero de 1998	789.095	12.5	1	1998	01/01/1998	31/01/1998	31	0.08	Ene	0.8
CHC	febrero de 1998	1785.304	12.5	2	1998	01/02/1998	28/02/1998	28	0.21	Feb	0.8
CHC	marzo de 1998	4867.428	12.5	3	1998	01/03/1998	31/03/1998	31	0.52	Mar	0.8
CHC	abril de 1998	6242.364	12.5	4	1998	01/04/1998	30/04/1998	30	0.69	Abr	0.8
CHC	mayo de 1998	4335.000	12.5	5	1998	01/05/1998	31/05/1998	31	0.47	May	0.8
CHC	junio de 1998	4189.000	12.5	6	1998	01/06/1998	30/06/1998	30	0.47	Jun	0.8
CHC	julio de 1998	2700.134	12.5	7	1998	01/07/1998	31/07/1998	31	0.18	Jul	0.8
CHC	agosto de 1998	5957.740	12.5	8	1998	01/08/1998	31/08/1998	31	0.64	Ago	0.8
CHC	setiembre de 1998	5063.078	12.5	9	1998	01/09/1998	30/09/1998	30	0.56	Set	0.8
CHC	octubre de 1998	4987.556	12.5	10	1998	01/10/1998	31/10/1998	31	0.54	Oct	0.8
CHC	noviembre de 1998	4484.123	12.5	11	1998	01/11/1998	30/11/1998	30	0.50	Nov	0.8
CHC	diciembre de 1998	5160.406	12.5	12	1998	01/12/1998	31/12/1998	31	0.55	Dic	0.8
CHC	enero de 1999	2289.976	12.5	1	1999	01/01/1999	31/01/1999	31	0.25	Ene	0.8
CHC	febrero de 1999	5417.760	12.5	2	1999	01/02/1999	28/02/1999	28	0.64	Feb	0.8
CHC	marzo de 1999	7851.194	12.5	3	1999	01/03/1999	31/03/1999	31	0.84	Mar	0.8
CHC	abril de 1999	8049.834	12.5	4	1999	01/04/1999	30/04/1999	30	0.89	Abr	0.8
CHC	mayo de 1999	8540.635	12.5	5	1999	01/05/1999	31/05/1999	31	0.92	May	0.8
CHC	junio de 1999	8206.122	12.5	6	1999	01/06/1999	30/06/1999	30	0.92	Jun	0.8
CHC	julio de 1999	7053.084	12.5	7	1999	01/07/1999	31/07/1999	31	0.76	Jul	0.8
CHC	agosto de 1999	4575.366	12.5	8	1999	01/08/1999	31/08/1999	31	0.49	Ago	0.8
CHC	setiembre de 1999	3853.260	12.5	9	1999	01/09/1999	30/09/1999	30	0.43	Set	0.8
CHC	octubre de 1999	2309.214	12.5	10	1999	01/10/1999	31/10/1999	31	0.25	Oct	0.8
CHC	noviembre de 1999	2023.390	12.5	11	1999	01/11/1999	30/11/1999	30	0.11	Nov	0.8
CHC	diciembre de 1999	1515.788	12.5	12	1999	01/12/1999	31/12/1999	31	0.16	Dic	0.8
CHC	enero de 2000	5287.885	12.5	1	2000	01/01/2000	31/01/2000	31	0.57	Ene	0.8
CHC	febrero de 2000	5122.494	12.5	2	2000	01/02/2000	29/02/2000	29	0.59	Feb	0.8

Fig. 3.Data loaded after cleaning, ordering, transformations and calculations.

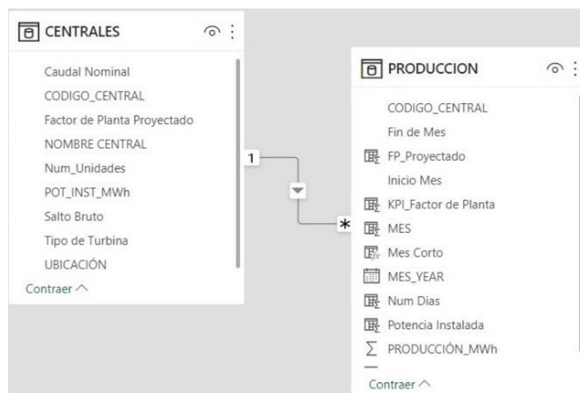

Fig. 4. Data modeling.

Step Four: Modeling.

The most relevant model is the predictive model. For this, models embedded in the Power BI platform were used. The exponential smooth, ARIMA, and one based on a Neural Network models were used. The graphical results can be seen in Fig. 5, Fig. 6 and Fig. 7.

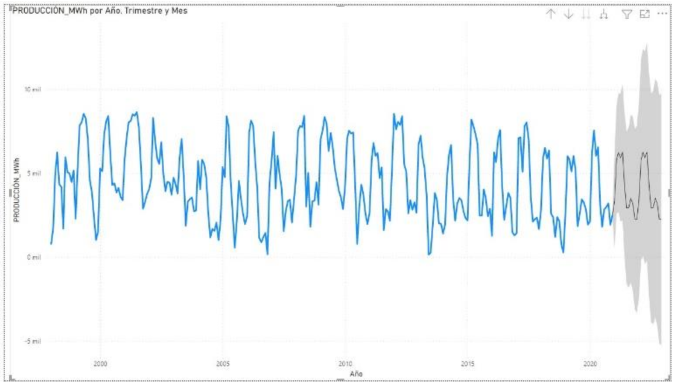

Fig. 5 Prediction with Exponenial Smooth Model

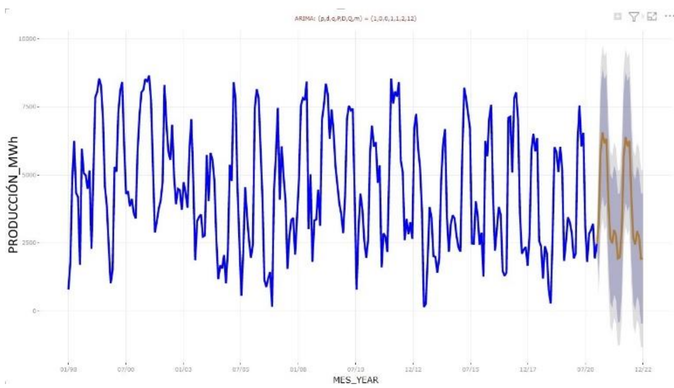

Fig. 6. Prediction with ARIMA Model

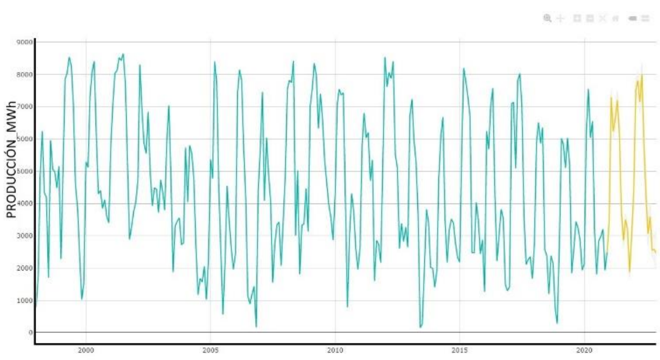

Fig. 7. Prediction with Neural Network model.

Taking into account that the objective of this work is to test the methodology in a specific case, we do not consider the explanation and mathematical deepening of the models necessary.

Step five: Evaluation.

This step presents the possibility of confusion. Really, the evaluation that the method requires is the evaluation facing the objectives of the business or data mining, not specifically of the models. However, in this specific case and many others, this evaluation can cover both aspects, as in this case where the best model is at the same time the one that best responds to the data mining objectives, that is, the prediction of the production of the next year.

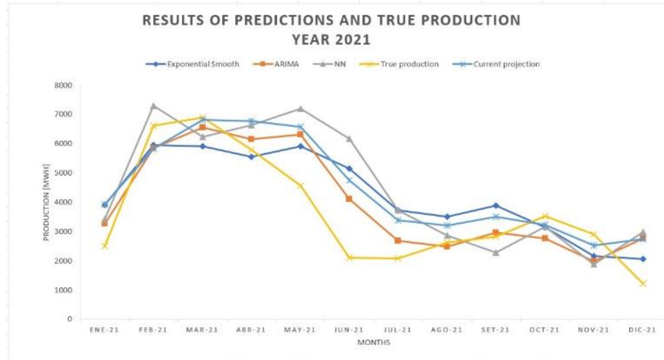

Fig. 8. Graphic results of the models.

The evaluation is carried out by comparing the results of each model with the actual production of the following year. We have generated the model with known data from 1998 to 2020 and compared the model with the true production in the year 2021. This comparison can be seen in Fig. 8. The numerical results can be seen in Table III.

TABLE III
RESULTS OF THE MODELS

Month	Smooth Exp.	ARIMA	NN	True production	Current projec.
Ene-21	3902	3265	3427	2508	3920
Feb-21	5952	5848	7297	6625	5837
Mar-21	5916	6560	6231	6908	6821
Abr-21	5562	6156	6640	5787	6780
May-21	5917	6327	7205	4577	6587
Jun-21	5145	4118	6182	2109	4742
Jul-21	3719	2684	3725	2082	3387
Ago-21	3515	2490	2856	2632	3202
Set-21	3894	2976	2280	2828	3516
Oct-21	3141	2774	3169	3521	3223
Nov-21	2154	1981	1878	2910	2525
Dic-21	2071	2774	2992	1221	2740

Finally, we calculated the Mean Square Error for all three models. The results of this calculation can be seen in Table IV.

TABLE III
MSE ERROR

Mean Square Error for the models				
Error	***SE Model***	***ARIMA Model***	***NN Model***	***Mean Projection***
MSE	1706240	1063938	2779211	1635086

With the criteria of the MSE, it is determined that the ARIMA model has the best performance.

Step six: Implementation.

The implementation happens, in this case, to present annually or month by month according to the criteria of the company these projections.

The relevant thing in this case is that being developed in a BI platform, as in this case Power BI, allows a connection to the update file of the production reports that will allow a manual or automatic update with which the model will be recalculated obtaining a new projection.

III. DISCUSSION OF RESULTS

The results of the application of the methodology show:

That the methodology can be followed for the problem posed in the case study and therefore is valid for this case.

The prediction models embedded in the BI platform generate a better prediction than the method followed by the company.

The parameter setting of the models is not so intuitive and therefore can be improved.

In fact, better prediction models can be generated, but this was not the objective of the work.

Considering that the experiments have been carried out with commercial and easily accessible hardware and software, the solution and methodology can be scaled to SMEs.

Knowledge of prediction models, use of the BI platform, installation of models, basic DAX have been required; Therefore, a limitation may be having minimally trained personnel for these actions.

IV. CONCLUSIONS

The present work shows that the use of a data mining methodology in a business intelligence environment can obtain useful results for an electricity generation SME.

The results do not require sophisticated software or hardware, so they are within the reach of SMEs.

Although the goal of this work is not to create an optimal predictive model, the case study demonstrates that with default adjustments, the tested models deliver better outcomes compared to the conventional practices used in the analyzed company.

Finally, it is shown that when the objectives of data mining are clear and defined, the reference model used (CRISP-DM) is still valid.

REFERENCES

[1] O. R. Florian Castillo, A. D. Aguilar Rivas, and F. G. Limo Cisterna, "Commercial Strategies for Customer Loyalty of an SME Company in the Commercial Sector," *Latin American and Caribbean Consortium of Engineering Institutions*, Aug. 2022, doi: 10.18687/LACCEI2022.1.1.168.

[2] G. Srivastava, M. S, R. Venkataraman, K. V, and P. N, "A review of the state of the art in business intelligence software," *Enterp Inf Syst*, vol. 16, no. 1, pp. 1–28, Jan. 2022, doi: 10.1080/17517575.2021.1872107.

[3] J. Boyer, B. Green, and B. Frank, *Business Intelligence Strategy: A Practical Guide for Achieving BI Excellence*. MC PR LLC, 2010.

[4] M. Muntean, D. Dănăiață, L. Hurbean, and C. Jude, "A Business Intelligence & Analytics Framework for Clean and Affordable Energy Data Analysis," *Sustainability*, vol. 13, no. 2, p. 638, Jan. 2021, doi: 10.3390/su13020638.

[5] G. Schuh *et al.*, "Data Mining Definitions and Applications for the Management of Production Complexity," *Procedia CIRP*, vol. 81, pp. 874–879, 2019, doi: 10.1016/j.procir.2019.03.217.

[6] O. D. Castrillón Gómez, J. A. Arango Marin, and J. A. Giraldo Garcia, "Data Mining in Labor Productivity," *Latin American and Caribbean Consortium of Engineering Institutions*, no. 2, Aug. 2022, doi: 10.18687/LACCEI2022.1.1.19.

[7] M. Molina-Calderon *et al.*, "Application of Machine Learning for the Prediction of Covid19 Through Classification Techniques and Supervised Learning," *Latin American and Caribbean Consortium of Engineering Institutions*, Aug. 2022, doi: 10.18687/LACCEI2022.1.1.425.

[8] P. M. Tocto Inga, G. T. Huamaní Huamaní, and E. F. Villacorta Casma, "Application of artificial intelligence in the management of a public university in Peru: A case of supervised machine learning using neural networks to classify whether and engineering students would graduated in 5 years," *Latin American and Caribbean Consortium of Engineering Institutions*, Aug. 2022, doi: 10.18687/LACCEI2022.1.1.565.

[9] C. Schröer, F. Kruse, and J. M. Gómez, "A Systematic Literature Review on Applying CRISP-DM Process Model," *Procedia Comput Sci*, vol. 181, pp. 526–534, Jan. 2021, doi: 10.1016/J.PROCS.2021.01.199.

[10] F. Medina Rojas, master Telemática, H. Arguello Fuentes, C. Engineering, and C. Gómez Santa Maria, "Diseño de Patrones de Estratificación de Cultivos en el Departamento del Huila Colombia, Minería de Datos para una Agricultura de Precisión," *Latin American and Caribbean Consortium of Engineering Institutions*, Aug. 2016, doi: 10.18687/LACCEI2016.1.1.104.

[11] K. Villamizar Suaza, D. Soto Durán, J. C. Giraldo Mejía, and J. Jimenez Builes, "Modelo de pruebas en proyectos BI," *Latin American and Caribbean Consortium of Engineering Institutions*, Aug. 2016, doi: 10.18687/LACCEI2016.1.1.271.

[12] A. B. Coronel-Alayo, P. G. Castillo-Domínguez, and V. E. Davila-Rodriguez, "Sistema De Inteligencia De Negocios En La Administración De Cultivos Para La Producción Agrícola," *Latin American and Caribbean Consortium of Engineering Institutions*, Aug. 2021, doi: 10.18687/LACCEI2021.1.1.262.

[13] O. D. C. Gomez, J. A. G. Garcia, and J. A. A. Marin, "Técnicas De Minería De Datos En La Industria Automotriz," *Latin American and Caribbean Consortium of Engineering Institutions*, Aug. 2021, doi: 10.18687/LACCEI2021.1.1.47.

[14] J. D. M. Bastidas, A. P. Gurgos, M. Garza, Á. Galindo, and S. Obregón, "Diseño E Implementación De Modelos De Pronósticos Para Estimar La Demanda De Bebidas En Tiendas De Conveniencia," *Latin American and Caribbean Consortium of Engineering Institutions*, Aug. 2021, doi: 10.18687/LACCEI2021.1.1.444.

[15] H. Wiemer, L. Drowatzky, and S. Ihlenfeldt, "Data Mining Methodology for Engineering Applications (DMME)—A Holistic Extension to the CRISP-DM Model," *Applied Sciences 2019, Vol. 9, Page 2407*, vol. 9, no. 12, p. 2407, Jun. 2019, doi: 10.3390/APP9122407.

[16] S. Studer *et al.*, "Towards CRISP-ML(Q): A Machine Learning Process Model with Quality Assurance Methodology," *Machine Learning and Knowledge Extraction 2021, Vol. 3, Pages 392-413*, vol. 3, no. 2, pp. 392–413, Apr. 2021, doi: 10.3390/MAKE3020020.

[17] S. A. Mohd Selamat, S. Prakoonwit, R. Sahandi, W. Khan, and M. Ramachandran, "Big data analytics—A review of data-mining models for small and medium enterprises in the transportation sector," *Wiley Interdiscip Rev Data Min Knowl Discov*, vol. 8, no. 3, p. e1238, May 2018, doi: 10.1002/WIDM.1238.

[18] Huang Lei, Huang Yifei, and Guo Yi, “The research of business intelligence system based on data mining,” in *2015 International Conference on Logistics, Informatics and Service Sciences (LISS)*, Jul. 2015, pp. 1–5. doi: 10.1109/LISS.2015.7369786.

[19] C. Shearer, *The CRISP-DM Model: The New Blueprint for Data Mining*. 2000.

[20] M. Ali *et al.*, “A Data-Driven Knowledge Acquisition System: An End-to-End Knowledge Engineering Process for Generating Production Rules,” *IEEE Access*, vol. 6, pp. 15587–15607, 2018, doi: 10.1109/ACCESS.2018.2817022.

[21] S. Huber, H. Wiemer, D. Schneider, and S. Ihlenfeldt, “DMME: Data mining methodology for engineering applications – a holistic extension to the CRISP-DM model,” *Procedia CIRP*, vol. 79, pp. 403–408, 2019, doi: 10.1016/j.procir.2019.02.106.

[22] F. Martinez-Plumed *et al.*, “CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories,” *IEEE Trans Knowl Data Eng*, vol. 33, no. 8, pp. 3048–3061, Aug. 2021, doi: 10.1109/TKDE.2019.2962680.

[23] H. A. Ayad, L. A. Al-Kilani, R. Arshad, M. A. Al-Obadi, H. T. Hussein, and M. Kucukvar, “Developing an Interactive Data Visualization Platform to Present the Adaption of Electrical Vehicles in Washington, California and New York,” in *2020 IEEE 7th International Conference on Industrial Engineering and Applications (ICIEA)*, Apr. 2020, pp. 731–737. doi: 10.1109/ICIEA49774.2020.9101928.

[24] J. M. Palma-Ruiz, A. Torres-Toukoumidis, S. E. González-Moreno, and H. G. Valles-Baca, “An overview of the gaming industry across nations: using analytics with power BI to forecast and identify key influencers,” *Heliyon*, vol. 8, no. 2, p. e08959, Feb. 2022, doi: 10.1016/j.heliyon.2022.e08959.

[25] F. Pietrapertosa *et al.*, “An educational awareness program to reduce energy consumption in schools,” *J Clean Prod*, vol. 278, p. 123949, Jan. 2021, doi: 10.1016/j.jclepro.2020.123949.