Survey into predictive key performance indicator analysis from data mining perspective

Akshay Thakur
Productivity Controlling & Production system
Volkswagen AG
Wolfsburg, Germany
akshay.vivek.thakur@volkswagen.de

Robert Beck
Productivity Controlling
Volkswagen AG
Wolfsburg, Germany
robert.beck3@volkswagen.de

Sanaz Mostaghim
Faculty of Computer Science
Otto von Guericke University
Magdeburg, Germany
sanaz.mostaghim@ovgu.de

Daniel Großmann
Faculty of Engineering & Management
Technische Hochschule Ingolstadt
Ingolstadt, Germany
Daniel.Grossmann@thi.de

***Abstract*—Predictive analytics is seen as one of the emerging technology in this digital age of big data. Computational processing power and speed has grown exponentially in the last few years that has made predictive analytic practical for application in different organization. Manufacturing industries has huge amount of data in different shapes and forms, and keep regular track of their performance by monitoring key performance indicators defined under business strategy. Prioritizing and predicting these key performance indicators provide organization cutting edge as compared to competitors by being proactive rather than reactive. As compared to traditional business intelligence tools where focus is on static report or dashboards about past data, predictive analysis focuses on estimating outcomes with the objective of driving better business performance. Moreover, it is also being adopted for decision-making tools. Different data mining techniques are applied in the field of performance management system as per individual or project need. Many researches has developed different ideas to understand and evaluate complex intervened key performance indicator relationships in performance measurement system. The aim of the paper is to present comprehensive version of predictive key performance indicator analysis from its background to state of the art, describing various data mining standards, methodologies as well as industrial and research application. The paper also studies various surveys regarding predictive analytic for business application to identify different best practices in this field.**

***Index Terms*—KPI selection, KPI relationship, key performance indicators, predictive analysis, survey, data mining, best practices**

I. Introduction

Performance measurement system are necessary for companies or manufacturing firms to understand the state of the production system/network, which are measured in term of Key Performance Indicator (KPI). These are numbers/values, which are compared and analyzed against defined internal targets or an external benchmark from the competitors to indicate performance of the system. KPI helps in understanding and improving manufacturing performance, by eliminating wastes from the perspective of Lean and achieving strategic goals from corporate perspective [1].

KPI selection and evaluation are one of the crucial steps of the modern industry, as setting up of inadequate objectives or error in judgement by middle or top management will compound themselves throughout the entire organization [2]. Traditional KPI prediction uses forecasting with aid of stochastic statistical models which lacks the ability to identify intricate relationships between them. In many manufacturing industries, prioritizing of KPI has become bottleneck in their effort of improving performance management system. Analytical hierarchy process is applied as decision-making tool for assigning relative weights to KPI using pairwise comparison. It does not specify relationship between KPI and their individual impact on complete performance system [3] [4]. Predictive analytics in conjunction with data mining provides insight into massive amount of data, whereas traditional analysis only leads to information discovery with techniques like descriptive statistics, dimensional slicing and hypothesis testing [3]. Predictive analytics provide opportunity to derive value from their data by considering it as valuable asset. It refers to tools and techniques from the field of statistics, computational science and other quantitative disciplines used to determine probability of futuristic outcomes using past information. Organizations used or want to use predictive analytic for vast spectrum of applications like road safety relevant key performance indicator selection, diseases prediction, sales prediction, market analysis, customer segmentation to name a few. Wide range of data like structured data, demographic data, time series data, real time event data is being used for predictive analytic by extracting insight from data. Predictive analytics with advance visualization tools and open source technologies has given solid foundation for fast moving landscape. Popular predictive analytics technique include data regression, classification and clustering with various use cases like financial fraud detection, risk analysis, employee churn prediction, customer related analytics, predictive maintenance and many more. Survey revealed that respondents using predictive analytics in industry had 8% increased chance in success with their analytics programs

978-1-7281-8956-7/20/$31.00 ©2020 IEEE

compared to others. Companies using data analytics along with big data were found to be having 5-6% more productivity rates and profitability as compared to others. Adoption of predictive analytics in businesses is increased from 21% in 2007 to 40% in 2016 which proves it as emerging technology [5] [6].

As virtue of different data mining techniques being used for enhancing various features of KPI, this paper focuses on surveying predictive KPI analysis along with their industrial application and applied methodologies. Scope of the survey paper is limited to research paper published after year 2000 and constrained to data mining application for KPI. This paper is organized as follows. In Section II, KPI background with different KPI types, characteristics and visualization is discussed. Section III focuses on standards for predictive modeling. Various data mining models and applications for KPI selection, KPI prediction and performance measurement system is reported in Section IV. Section V covers best practices and current trends in predictive analysis practical application. Finally, Section VI concludes the paper.

II. KPI Background

A. Defining KPI

International Organization for Standardization (ISO) has defined different standards regarding KPI, namely ISO 22400 for manufacturing operations management, ISO 14031 for environmental performance evaluation and ISO 13053 for defining Six Sigma performance improvement methodology. This survey will focus majorly on KPI defined under ISO 22400. These standards are defined for "Automation System and Integration - KPI for manufacturing operations management" [7]. Manufacturing operations Management includes four major categories of operation management: production, maintenance, quality, and inventory; under which 34 KPIs are identified and defined. Few examples of these KPIs are: Allocation Efficiency, Availability, Quality Ratio, Scrap Ratio, Utilization Efficiency etc [8]. The ISO standards defines KPI by means of their formula and corresponding elements, their units/dimensions, their time behaviors and other characteristics as shown in Table I. This standardized schematic helps in providing a means to categorize productivity tools that are inter-operable as well as used across different applications [1].

Doran [2] proposed the SMART philosophy to compose the Management's goals and objectives by considering them as critical step in company's management process. The SMART represents:
Specific - target a specific area for development
Measurable - quantify/indicator for improvement
Assignable - responsible personnel/team
Realistic - objectives that can be achieved with available resources
Time related - time frame for result to be achieved.

B. Characteristics of KPI

Based on extensive research during KPI workshops in public and private sectors, Parmenter [9] defined below characteristics for KPI

TABLE I
Structure of KPI description [8]

Content	KPI Description
Name	Name of the KPI
Description	Brief description of KPI
Unit of measure	Unit/dimension in which KPI is expressed
Production Methodology	Discrete/Batch/Continuous
Timing	Real time/On demand/Periodically
Audience	Operators/Supervisors/Management
Effect Model Diagram	Graphical representation of all the influencing elements for the KPI
ID	A user defined unique identification of the KPI in the user environment
Scope	Identification of KPI relevant element, which can be work unit, personnel, product or work center
Formula	Mathematical formula specified in terms of elements
Range	Specifies the upper and lower logical limits of KPI
Trend	The direction of improvement, lower is better/more is better

- Non-Financial: KPI should not be defined in terms of money, but rather on influencing elements.
- Monitored: They should be measured frequently (24/7, daily, weekly).
- Top Management Focus: Acted upon by Chief Executive Officer and senior management.
- Simple: KPI as well as relevant corrective action should be easy to understand.
- Team Based: Responsibility to be assigned to the team members who are working closely together.
- Significant Impact: KPI should have significant and positive impact on organization and needs to be tested beforehand to avoid deteriorating influence.

Difference between KPI and KRI (Key Result Indicator) lies in its time-frame outlook. KRI focuses on outcome measures, which typically are past activity over months and quarters. KPI are future oriented measures, which focuses on specific activity.

C. Types of KPI

KPI are generally classified into leading and lagging KPI. Leading KPI measure activities, which has significant impact on future performances, whereas lagging indicator measures output of past events [10]. Another type of KPI classification is into below categories (as shown in Table II) based on industrial needs and requirements of multiple production levels [7].

D. KPI Monitoring & Visualization

KPI monitoring ensures that complete system works at optimum efficiency and at desired quality standards [7]. Manufacturing organizations aim to integrate their business functions with digital system with the help of enterprise database. These digital system that are based on Computer

TABLE II
KPI CLASSIFICATION WITH EXAMPLES [7]

KPI Types	KPI Example
Supplier	Supplier cost, Supplier delivery, Supplier utilization
Sustainability	Eco-friendly, Raw material stock, Re-utilization
Design	Design effort, frugal product innovation index, Data flexibility during project
Manufacturing	Lead time, Production rate, Mean time to repair, Quality, Purchasing cost, Total production time
Customer	Customer satisfaction, Customer acceptance, Customer involvement
Financial	Production cost, Energy cost, Product cost, Transportation cost

Integrated Manufacturing (CIM) assist in KPI monitoring. KPI monitoring can be online monitoring (with real time data) and offline monitoring, which can be done either by humans, electronic sensors or by software tools based on simulation environments [7] [11].

KPI Visualization helps in visualizing the suitable metrics accurately to assist the decision makers. Different visualization methods are dashboards, bullet graphs, 3D scatter plot, slider KPI charts, pie charts, traffic light KPI charts, speedometer KPI charts etc. [7]. Currently, these visualizations are core-integrated part of various Business Intelligence tools. Elliot [12] presented Joint Medical Asset repository (JMAR), a relational database system to integrate KPI like medical supply and medical maintenance management systems.

E. Manufacturing KPI example

Overall equipment effectiveness (OEE) is one of the most common KPI used in manufacturing industry to track manufacturing productivity. OEE is defined as measure of total equipment performance based on availability (A), quality (Q) and performance (P) of the output [13]. From below formulas (See 1-4), it is clear that KPI in Performance Management system are intervene and are dependent on many influencing parameters. OEE also indirectly affect other parameters like investment, incurred costs, energy consumption etc. which cannot be explicitly seen in the below formulas [14].

$$OEE = \text{Availability} * \text{Performance} * \text{Quality} \tag{1}$$

$$Availability = \frac{\text{Actual Operating time}}{\text{Planned production time}} \tag{2}$$

$$Performance = \frac{\text{Actual production rate}}{\text{Ideal production rate}} \tag{3}$$

$$Quality = \frac{\text{Good parts}}{\text{Total parts}} \tag{4}$$

In [14], author defines OEE as multi objective optimization to perform multiple criteria decision making. In addition, comparison between OEE as single objective and multi objective is done for determining optimization performance on steel production machine. In [13], OEE simulation model is developed based on fuzzy model and data mining by taking in account other influencing factors. In [15], data mining clustering alongside association rule algorithm is implemented to analyze key factors affecting OEE. It is clear from the above stated research for OEE optimization application that, key performance indicators are complex entities and have diverse techniques for predictive modelling. This paper by studying various predictive KPI application attempts to map out different significant data mining methodologies.

III. PREDICTIVE MODELING STANDARDS

As opposed to descriptive statistics (standard deviation, mean, mode and histogram) used for forecasting, predictive analytics uses computing power to identify meaningful patterns and relationships, with the use of machine learning, neural computing, computational mathematics and artificial intelligence techniques. In 1996, several data mining industry players created standard methodology named as "Cross Industry standard Process for Data mining (CRISP-DM)" for predictive modeling. As per 2007 survey (167 respondents) who have implemented predictive analyses, only 15% used CRISP-DM and majority were found using own methodology (52%) or vendor methodology (29%). CRISP-DM define process for data mining as below:

- **Project definition:** It comprises of determining business objectives along with data mining goal and producing project plans.
- **Data understanding:** Data quality is verified after initial data collection, data description and data exploration. Report is generated for each mentioned steps together with exploratory analysis report.
- **Data preparation:** Data for further analysis is selected/excluded based on data mining goals, technical constraints or data types. Data cleaning, data preparation, data integration and data transformation are crucial steps that are involved.
- **Modeling:** Create, test and validate models on basis of accuracy and generality of the model. Supervised data mining projects like classification are tested based on error rates calculation. In case of multiple models, models are ranked as per evaluation criteria.
- **Evaluation:** Model is evaluated to assess its fulfillment level of business requirement.
- **Deployment:** Strategy for deployment is identified along with model monitoring and maintenance plan in future [16] [17].

Similar to CRISP, there are various framework like SEMMA (Sample, Explore, Modify, Model & Assess) developed by SAS Institute, DMAIC (Define, Measure, Analyze, Improve & Control) and KDD - Knowledge discovery in databases (Selection, Preprocessing, Transformation, Data mining, Interpretation & Evaluation) [3].

IV. PREDICTIVE ANALYTICS FOR KPI

Predictive modeling is classified into two categories, i.e., data based method and model based method. Model based method entrust on process mechanism and knowledge. On the other hand, data base model has no such restriction and can be applied to different manufacturing processes. Various data based predictive modeling methods include random forest, artificial neuron network, support vector regression [18]. Predictive modeling differ from data reporting, analysis and monitoring with its ability to deliver high business value at more complex level (See Fig. 1).

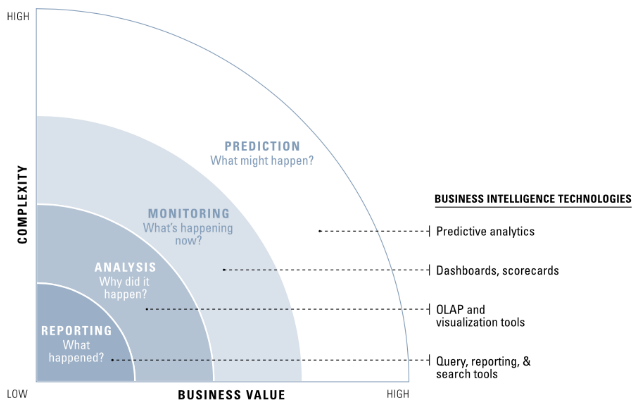

Fig. 1. Spectrum of Business Intelligence technologies [16].

Another predictive modeling classification is based on statistical model and artificial intelligence models. Artificial intelligence models consists of Fuzzy logic, Neural networks, Genetic algorithm, Rule Induction, Principal component analysis, Kohonen Network, Nearest neighboring pairing etc. Statistical model requires lots of historical data. Linear Regression, logistic Regression, Nova, Time series forecasting, Survival analysis, non-linear regression etc. are example of statistical models [19]. For decimal, integer, duration, time, date-time, and other date type of metrics, regression is used to create a prediction model, while for Boolean and string types of metrics, a classification models such as decision trees are used to create predictive models. In [20] , event driven KPI prediction was proposed using approaches like time series KPI prediction and metric aggregation for Quality of Service Management. The selection of data mining algorithm depends on project objective, type and structure of data set, available computational power, outliers presence and so on [3].

A. Balance Score card & Performance Management System

Performance measurement system (PMS) is defined as the process of quantifying the efficiency, effectiveness, relevance and effectivity of action, which is done by using set of metrics/KPI. The PMS system is examined at three different levels: individual performance measure, PMS as an entity and relationship between PMS and environment within which it operated [21] [22]. Kaplan and Norton presented Balance score card (BSC) with four perspective, which needs to be balanced in performance measurement: financial, customer, internal business process and development perspective. BSC with its lagging (financial) and leading (customer, business process & development) consideration proves to be strategic management system rather than operational [23]. BSC overcomes the problem of strategic alignment and balance approach for identifying the most appropriate KPI. Due to presence of various individual KPI in BSC, selection and trade off in obtaining perfect strategy is one of its limitation. The study of cause and effect of one KPI on another is also complex in this convoluted relationship [4]. Fuzzy Analytical hierarchical process is used extensively for evaluation of performance measurement system because of its capability to merge qualitative and quantitative information for decision-making. It integrates the concept of fuzzy set theory and hierarchical structure analysis. Different data mining applications related to BSC and PMS are listed in Table IV.

B. Application of KPI Selection

In [24], the authors focused on constructing three sets of road safety performance indicator from the list of 30,33 & 20 performance indicator for regional road, urban road and highway respectively. The importance of the performance indicators was identified by applying Fuzzy Delphi method and Grey Delphi method. The Fuzzy Delphi method was chosen over Delphi method due to its less expert data requirement resulting in low cost as well as results obtained are reasonable. The analysis showed that number of performance indicator identified using Fuzzy Delhi method were more than Grey Delphi method. Intersection of both results was done to obtain final set of performance indicators. The final calculated performance indicator were found to be feasible and practical for implementation. Analogue to this, in [25], authors proposed hierarchical fuzzy TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) along with experts knowledge to combine 11 layered performance indicator to find one overall performance index. The result comparison of Fuzzy TOPSIS with hierarchical fuzzy TOPSIS showed that the later model is more sensitive to indicator structure change, proving its robustness.

In [26], developed a method to find suitable KPI to associate with business objectives in semi-automated way for online education courses. Modified data mining process consisted of five processes: preprocessing, outliner's detection, difference series calculation, analysis of pair wise relation between series and analysis of compound relationship. Visualization technique is used along with data mining process due to difficult interpretation of large amount of data. Data was analyzed using Support vector machines, Random forest and Multilayer perceptron network, which provided similar results with 80%-84% accuracy. Based on final weightage to the KPI, single crucial KPI was identified using Random forest method. In [27], case study for Dutch ERP (Enterprise resource Planning) software vendor was carried to identify relevant KPI from list of 19 KPI. By means of the automatically determined relevance factors of KPIs, the prediction of relevant KPIs for organizations is done which makes this method unique. The prediction modeling technique used were Logistic regression, SVM (Support vector machines), Decision Tree, Random forest and stacked (Decision Tree & SVM). 64% & 74%

balanced accuracy was achieved at predicting the relevance of 13 KPI & 6 KPI respectively. The future work aims on designing engaging dashboard with automatic relevant KPI identification using decision graph.

In [28], the study was done for textile industry to compare 73 ERP packages to select suitable one, based on KPI fulfillment by using Fuzzy AHP (Analytic hierarchy process) and balance score card. Fuzzy AHP was favored over normal AHP method because of its ability to tackle decision making in complex environments and avoid vagueness/ imprecision due to human interference. In [29], research was done to identify the fundamental supplier combination that will minimize the cost-quality KPI score. The model combined fuzzy theory, genetic algorithm and T- transformation technology. The fuzzy theory was used to quantify the KPI data, the T-transformation to deal with the defuzzied data for integrating the different attributes data such as cost and quality and genetic algorithm to search intelligently for the optimal part change solution. The result obtained by above model was compared to linear programming model, which proved that proposed model was reliable. In [30], fuzzy multiple-criteria decision making (FMCDM) is applied to adopt national quality award criteria as the Six-Sigma project selection criteria. The result indicated that FMCDM was robust and flexible for group decision-making process.

C. Applications of KPI Prediction

In [31], one of prominent KPI from automotive industry was predicted using historical data. CRISP-DM methodology was applied as is. Algorithm tested were Random forest, SVM, M5 (modified regression tree algorithm) and partial least square on 71 samples with 6 predictors. Prediction accuracy was verified using target Root Mean Square Error (RMSE) value and comparison with Naive prediction value. The RMSE target was 0.25 and SVM model was best performing model with 0.31 RMSE compared to other tested models. In [32], case study in telecommunications business unit was carried out to find the KPI correlation and their influence on other KPI on organizational level. Decision tree along with Rule extraction matrix identified the influencing factors among KPI.

In [33], case study was presented to predict the student dropout rate for open courses. CRISP-DM was used as it is. 32 KPI were used as initial features for data mining, which were later reduced to six after data pre-processing stage. Prediction process was found to be influenced by feature selection algorithm, which played pivotal role in prediction. Data model was tested on three different cases: (1) considering all attributes (2) considering top 10 attributes (3) applying data balancing algorithm on selected attributes and using this data for classification. Rule based classification like Jrip, NNge (Nearest neighbor like algorithm), conjunctive rule, DTNB (decision table/Naive Bayes hybrid classifier) and PART (partial decision tree algorithm) were considered. Different form of Decision tree like J48, NBTree, REPTree, SimpleCart were applied to identify relevant factors affecting dropout rate.

In [18], a distributed predictive modeling framework is proposed for prediction and diagnosis of KPI in plant wide processes. Furthermore, the framework included a diagnostic scheme for identification of KPI performance degradation. 800 data samples were used as training data set and 500 data samples were used as testing data-set, which were reduced to 160 & 100 respectively after multi-sample rate of input and output variables. Principal component analysis (PCA) was done at initial stage to extract information from each block of process. The retained PCA components were feed as input to Gaussian Process Regression models to identify significant process variable. In [34], decision tree were used to identify influential factors for KPI by combining process events and quality of service measurements. 31 influencing parameters were identified and processed using WEKA (Waikato Environment for Knowledge Analysis) toolkit with J48 and ADTree (alternating decision tree) algorithm. AD tree was found to be producing bigger tree as J48 for the same number of instances, and tree were found to get bigger with increase in number of instances for both cases. The case study identified that decision tree has risk of hiding influential factor due to multilevel dependencies between KPI. Another risk is that as decision tree gets bigger, various influencing factor comes into picture that has very marginal influence.

In [19], authors carried out survey on clinical prediction models for diabetes prediction. This paper showed that hybrid models produce more accuracy than traditional models. The hybrid model used is elastic net regression model, which is a combination of LASSO (Least Absolute Shrinkage and Selection Operator) and Ridged Regressions. The survey also pointed that there are many research gaps related to usage of large data sets, outliers detection and prediction model improvement. Summary of different industrial application for KPI prediction as well KPI selection can be seen in Table III.

V. Best Practices In Practical Predictive Analytics

Transforming Data with Intelligence (TDWI) conducted worldwide survey in quarter II/2018 to explore best practices regarding predictive modeling in organizations, which included analytics, architecture and corporate IT from wide variety of industries as survey respondents. From 544 correspondents, 505 were using predictive analysis (active group), 40% were planning to implement it in future (exploring group) and 10% had no plan for it. The report found out that active users applied predictive analysis in direct marketing (52%), retention analysis (52%), cross sell (49%) and default prediction (46%) mostly. Predictive maintenance and image classification were least used applications. The survey highlighted that structured, demographic and geospatial data was used more compared to IOT (Internet of Things) sensor data and external text data by active users. The trend signified that organization want and need to utilize divergent data since it widens the data set and add value to the prediction at times. It also highlighted the fact that being focused on Business Intelligence activities (like reports and dashboard) and lack

TABLE III
SUMMARY OF KPI SELECTION AND PREDICTION APPLICATION

Research Paper	Prediction model	No of KPI	CRISP-DM	Practical application	Application type	Data Mining task
Ref. [24]	Fuzzy Delphi method, Grey Delphi method	20-33	No	Road safety KPI	KPI Selection	Clustering
Ref. [25]	Hierarchical fuzzy TOPSIS	11 layers	No	Road safety KPI	KPI Selection	Multi criteria decision making
Ref. [26]	SVM, Decision Tree, Multilayer perceptron	14	Partly	Online Education course	KPI Selection	Classification
Ref. [27]	SVM, Random Forest, Decision Tree (DT), Logistic regression, Stacked (DT & SVM)	19	Partly	Enterprise Resource Planning	KPI Selection	Classification
Ref. [28]	Fuzzy AHP	21	No	Textile industry	KPI Selection	Multi criteria decision making
Ref. [29]	Fuzzy theory + T Transformation + Genetic Algorithm	16 (supplier combination)	No	LCD manufacturer	KPI Selection	Optimization problem
Ref. [30]	Delphi Fuzzy multi criteria decision making	33 (sub-criteria)	No	LCD manufacturer	KPI Selection	Multi criteria decision making
Ref. [31]	SVM , M5 Random Forest, Partial Least Square	6	Yes	Automotive industry	KPI Prediction	Regression
Ref. [32]	Decision tree + Rule extraction	35	Yes	Telecommunication business unit	KPI Prediction	Classification
Ref. [33]	Jrip, NNge, DNTB PART, Conjunctive rule	32	Yes	Dropout rate for open course	KPI Prediction	Rule based classification
Ref. [18]	Gaussian Process Regression	33	Partly	Tennessee Eastman Process	KPI Prediction	Regression

of machine learning skills apprehended the non-users group from using predictive analytics. Technologies with graphical user interface, workflow & versioning, collaborative features and persona-driven features were found alluring aspects for new users [5].For predictive analytics return on investment is not necessarily calculated in money but rather in its in value for assisting in better decisions making, a strong understanding of behavior, improved operational efficiencies or decrease risk [6].

Automated task are preferred over manual task in this digital era. 16% correspondent were found to be using automated tools for predictive analysis. Automation in predictive analysis procedure include automatic data preparation, model building and monitoring models. This automation in turn benefited other people to build predictive model (31%) and save time in creating model (24%). Predictive analysis was carried out widely with the help of open source tools like R (72%), Python (59%), and Spark among active users (244 respondents). 79% respondents stated that models are reviewed by an inside experts after built and 48% claimed that certain models can be only built by experts (219 active group respondents). Making predictive analysis as part of business process was established as top strategy for gaining value from it in organization.Analytics platform (like Data warehouse, Data Lake, Hadoop) in organization infrastructure and public cloud, are used currently and are also planned for the future. The survey emphasized that collaboration among inside and outside organization, result publication within organization and analyst collaboration were top best practices to help democratize analytics in an organization [5]. From the prescriptive analytics solutions perspective, vendors are categorized mainly into - full life cycle vendors, data science workbenches, automated tool vendors, application program interface (API) focused and those focused on intelligent solutions [6].

VI. CONCLUSION

The paper presents various methodologies for exploring different aspects of KPI like KPI selection, KPI prediction and evaluation of performance management system using data mining techniques along with their applications. From the survey, it is clear that KPI selection is preferably done using multi criteria decision making or by considering it as optimization problem. On the other hand, regression or classification based machine learning is preferred for KPI prediction based on the given problem (as shown in Table III). Reason behind limited academic research paper in data mining application for industrial use cases can be attributed to companies confidential KPI data.

Real time data analysis is seen as the future for business growth. To enable its success more research needs to be done in the field of standardization of model management after first implementation. There is also limited research found for use of structured data in conjunction with unstructured data (text analysis) for the KPI prediction and selection. Current research extensively focus mostly on data mining for enhancement of performance system. Future work for this paper is planned in two phases. First phase comprises of extending the survey to cover standards and practical application for data mining model management after implementation. Second phase will focus on practical side of the project in which different data mining methodologies are applied for KPI prediction and their performance is compared.

REFERENCES

[1] ISO/TC 184/SC 5, "Automation systems and integration — Key performance indicators (KPIs) for manufacturing operations management: Part 2 - Definitions and descriptions," International Organization for Standardization, Standard, Jan. 2014. [Online]. Available: https://www.iso.org/obp/ui/#iso:std:iso:22400:-2:ed-1:v1:en

[2] G. T. Doran, "There's a smart way to write management's goals and objectives." and miller," *Arthur F. & Cunningham, James A" How to avoid costly job mismatches" Management Review*, vol. 70, no. 11, pp. 35–36, 1981.

[3] V. Kotu and B. Deshpande, *Predictive analytics and data mining: concepts and practice with rapidminer*. Morgan Kaufmann, 2014.

[4] J. Cai, X. Liu, Z. Xiao, and J. Liu, "Improving supply chain performance management: A systematic approach to analyzing iterative KPI accomplishment," *Decision Support Systems*, vol. 46, no. 2, pp. 512–521, 2009.

[5] F. Halper, "Practical Predictive Analytics," TDWI, Tech. Rep., 2018.

[6] ——, "Tdwi navigator predictive analytics," TDWI, Tech. Rep., 2017.

[7] G. Lanza and D. Mourtzis, "Operational KPI evaluation framework," University of Patras, Patras, Tech. Rep., 2015.

[8] B. R. Ferrer, U. Muhammad, W. M. Mohammed, and J. L. Lastra, "Implementing and visualizing ISO 22400 key performance indicators for monitoring discrete manufacturing systems," *Machines*, vol. 6, no. 39, pp. 1–22, sep 2018.

[9] D. Parmenter, *Key Performance Indicators*, 3rd ed. New Jersey: John Wiley & Sons, Inc, 2015.

[10] M. Badawy, A. A. El-Aziz, A. M. Idress, H. Hefny, and S. Hossam, "A survey on exploring key performance indicators," *Future Computing and Informatics Journal*, vol. 1, no. 1-2, pp. 47–52, dec 2016.

[11] G. Chryssolouris, D. Mavrikios, N. Papakostas, D. Mourtzis, G. Michalos, and K. Georgoulias, "Digital manufacturing: history, perspectives, and outlook," *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, vol. 223, no. 5, pp. 451–462, 2009.

[12] E. B. Sloane, E. Rosow, J. Adam, and D. Shine, "JEDI-An Executive Dashboard and Decision Support System for Lean Global Military Medical Resource and Logistics Management," in *Proceedings of the 28th IEEE EMBS Annual International Conference*. New York City, USA: IEEE, 2006, pp. 5440–5443. [Online]. Available: www.esri.com

[13] M. G. R. Naik, V. Raikar, and P. G. Naik, "A simulation model for overall equipment effectiveness of a generic production line," *Journal of Mechanical and Civil Engineering (IOSR-JMCE)*, vol. 12, no. 5, pp. 52–63, 2015.

[14] A. Van Horenbeek, L. Pintelon, A. Bey-Temsamani, and A. Bartic, "Multi-objective optimization of oee (overall equipment effectiveness) regarding production speed and energy consumption," in *European conference of the prognostics & health management society. PHME*, vol. 14. Citeseer, 2014.

[15] L. Lin, H. Xiu, W. Yunfei, and W. Guangchao, "Research on overall equipment effectiveness for cigarette equipments based on data mart and data mining," *DEStech Transactions on Engineering and Technology Research*, no. apetc, 2017.

[16] W. W. Eckerson, "Predictive Analytics: Extending the value of your data warehousing investment," TDWI Research, Tech. Rep., 2007. [Online]. Available: www.tdwi.org

[17] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, R. Wirth *et al.*, "Crisp-dm 1.0: Step-by-step data mining guide," *SPSS inc*, vol. 9, pp. 1–73, 2000.

[18] Z. Ge, "Distributed predictive modeling framework for prediction and diagnosis of key performance index in plant-wide processes," *Journal of Process Control*, vol. 65, pp. 107–117, may 2018.

[19] N. Jayanthi, B. V. Babu, and N. S. Rao, "Survey on clinical prediction models for diabetes prediction," *Journal of Big Data*, vol. 4, no. 1, p. 26, 2017.

[20] L. Zeng, C. Lingenfelder, H. Lei, and H. Chang, "Event-driven quality of service prediction," in *International Conference on Service-Oriented Computing*. Springer, 2008, pp. 147–161.

[21] A. Neely, M. Gregory, and K. Platts, "Performance measurement system design: A literature review and research agenda," *International Journal of Operations and Production Management*, vol. 25, no. 12, pp. 1228–1263, 2005.

[22] M. S. Ravelomanantsoa, Y. Ducq, and B. Vallespir, "A state of the art and comparison of approaches for performance measurement systems definition and design," *International Journal of Production Research*, vol. 57, no. 15-16, pp. 5026–5046, 2019.

[23] I. Yüksel and M. Dağdeviren, "Using the fuzzy analytic network process (ANP) for Balanced Scorecard (BSC): A case study for a manufacturing firm," *Expert Systems with Applications*, vol. 37, no. 2, pp. 1270–1278, mar 2010.

[24] Z. Ma, C. Shao, S. Ma, and Z. Ye, "Constructing road safety performance indicators using fuzzy delphi method and grey delphi method," *Expert Systems with Applications*, vol. 38, no. 3, pp. 1509–1514, 2011.

[25] Q. Bao, D. Ruan, Y. Shen, E. Hermans, and D. Janssens, "Improved hierarchical fuzzy TOPSIS for road safety performance evaluation," *Knowledge-Based Systems*, vol. 32, pp. 84–90, aug 2012.

[26] J. Peral, A. Maté, and M. Marco, "Application of Data Mining techniques to identify relevant Key Performance Indicators," *Computer Standards and Interfaces*, vol. 54, no. 2, pp. 76–85, nov 2017.

[27] Ü. Aksu, D. M. Schunselaar, and H. A. Reijers, "Automated prediction of relevant key performance indicators for organizations," in *International Conference on Business Information Systems*. Springer, 2019, pp. 283–299.

[28] U. Cebeci, "Fuzzy AHP-based decision support system for selecting ERP systems in textile industry by using balanced scorecard," *Expert Systems with Applications*, vol. 36, no. 5, pp. 8900–8909, jul 2009.

[29] H. S. Wang and Z. H. Che, "An integrated model for supplier selection decisions in configuration changes," *Expert Systems with Applications*, vol. 32, no. 4, pp. 1132–1140, 2007.

[30] T. Yang and C. H. Hsieh, "Six-Sigma project selection using national quality award criteria and Delphi fuzzy multiple criteria decision-making method," *Expert Systems with Applications*, vol. 36, no. 4, pp. 7594–7603, 2009.

[31] S. A. P. de Oliveira, "Performance Management Analytics for the Automotive Industry," Master's thesis, Faculdade de Engenharia da Universidade do Porto, Portugal, 2015.

[32] R. Van Den Houten, "Proactive Business Intelligence," Master's thesis, Utrecht University, 2012.

[33] H. Gulati, "Predictive Analytics Using Data Mining Technique," in *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*. New Delhi, India: IEEE, 2015, pp. 713–716.

[34] B. Wetzsteina, P. Leitnerb, F. Rosenbergc, S. Dustdarb, and F. Leymann, "Identifying influential factors of business process performance using dependency analysis," *Enterprise Information Systems*, vol. 5, no. 1, pp. 79–98, feb 2011.

[35] E. K. Aydogan, "Performance measurement model for Turkish aviation firms using the rough-AHP and TOPSIS methods under fuzzy environment," *Expert Systems with Applications*, vol. 38, no. 4, pp. 3992–3998, 2011.

[36] Z. Cao, X. Liu, J. Hao, and M. Liu, "Simultaneous prediction for multiple key performance indicators in semiconductor wafer fabrication," *Chinese Journal of Electronics*, vol. 25, no. 6, pp. 1159–1165, 2016.

[37] C. Kahraman, T. Ertay, and G. Büyüközkan, "A fuzzy optimization model for QFD planning process using analytic network approach," *European Journal of Operational Research*, vol. 171, no. 2, pp. 390–411, jun 2006.

[38] A. M. A. El-mongy, A. E.-d. Hamouda, N. Nounou, and A. W. Abdelmoneim, "Design of prediction system for key performance indicators in balanced scorecard," *International Journal of Computer Applications*, vol. 72, no. 8, 2013.

[39] J. Daniel, R. M. Yusuff, and J. Jassbi, "The best selection of strategic plans in balanced scorecard using multi-objective decision making model," *African Journal of Business Management*, vol. 5, no. 3, pp. 681–686, 2011.

[40] A. H. Lee, W. C. Chen, and C. J. Chang, "A fuzzy AHP and BSC approach for evaluating performance of IT department in the manufacturing industry in Taiwan," *Expert Systems with Applications*, vol. 34, no. 1, pp. 96–107, 2008.

TABLE IV
DATA MINING APPLICATION FOR BALANCE SCORE CARD (BSC) AND PERFORMANCE MANAGEMENT SYSTEM (PMS)

Research Paper	Prediction model	Technique	Model comparison outcome	Number of KPI	Application
Ref. [35]	Rough AHP + Fuzzy TOPSIS	(1) Define the criteria for performance measurement (2) Calculate the weights of criteria with rough AHP (3) Evaluate the alternatives with fuzzy TOPSIS and determinate the final rank	No comparison made with other models	6 KPI (Effectiveness, efficiency, Occupational satisfaction, Decision, Risk, Quality)	PMS Application - Evaluation of performance measurement system of four Turkish aviation firm
Ref. [36]	Bayesian neural network + Bayes theorem	(1) Bayesian neural network (BNN) is used to establish the simultaneous prediction model for Multiple key performance indicators (MKPI) (2) Bayes theorem is used for model complexity controlling (3) BNN results are compared with Artificial neural network (ANN) and Selective naive Bayesian classifier (SNBC)	BNN outperformed ANN and SNBC for Multiple key performance indicator prediction	2 KPI (Cycle time and Equipment utilization)	PMS Application - Simultaneous Prediction for Multiple Key Performance Indicators in Semiconductor Wafer Fabrication
Ref. [37]	Fuzzy ANP + Fuzzy AHP	(1) Determine important degrees of involved KPI by using FAHP (Fuzzy Analytic Hierarchy Process) (2) Determining inner dependency matrix and interdependent priorities for both KPI (3) Determining overall priority of finals selected KPI using fuzzy ANP (Analytic Network Process)	Fuzzy AHP & ANP is compared with CRISP AHP & ANP	2 primary KPI (Technical product and customer needs requirements) and 12 secondary KPI	BSC Application - Quality deployment process (QFD) planning process for Turkish company producing window and door system
Ref. [38]	Fuzzy logic with neuron network	(1) Relations between KPIs identified using association rules (2) Fuzzy logic and ANN to predict KPI values laterally (3) Prediction results fused using C4.5 Decision tree	Error Rate: Fuzzy (6.3%), Neural network(5.2%), Fusion model(1.5%)	6 KPI Related to finance, customer, operations and learning & growth	BSC Application - Categorizing and predicting KPI for corporate balance score card in Egyptian company
Ref. [23]	Fuzzy ANP	(1) Determination of BSC perspective and related KPI (2) Structure the ANP model hierarchically and determine local weights of strategies (3) Fuzzy scales used to find inner dependency matrix (4) Calculate performance of organization based on fuzzy evaluation after assigning global weights for KPI	No comparison made with other models	4 primary KPI (financial, customer, internal business process and learning and development) and 16 secondary KPI	BSC Application - Application of BSC ANP model to measure performance in Turkish manufacturing firm
Ref. [39]	Fuzzy ANP	(1) Modeling of BSC based on input data (2) Objective weight identified by expert opinion (3) Gaining the final measure of objectives using the geometrical average (4) Best selection of strategic plan by Multi-objective decision making (Goal Programming)	No comparison made with other models	4 primary KPI (financial, customer, internal business process and learning and development) and 16 secondary KPI	BSC Application - Case study done in electronic and computer research center of the university for selection of strategic plan in BSC using goal programming model (type of multi objective decision making model)
Ref. [40]	Fuzzy AHP	(1) Expert feedback for BSC taken using conventional AHP questionnaire format (2) Feedback analyzed using designed FAHP program to obtain relative KPI importance	No comparison made with other models	4 primary KPI (financial, customer, internal business process and learning and development) and 14 secondary KPI	BSC Application - Proposed approach based on Fuzzy AHP and BSC for evaluating performance of IT department in manufacturing firm in Taiwan