The role of predictive data analytics in retailing

MOHAMMED JUNED SHAIKH SHABBIR
ME Student, Department of Computer Science and Engineering S.S.G.M.C.E
Shegaon, Maharashtra, India

C. M. MANKAR
Professor, Department of Computer Science and Engineering S.S.G.M.C.E
Shegaon, Maharashtra, India.

Abstract– big data analytics is the new practice in Business Analytics today. However, recent industrial surveys find that big data analytics may fail to meet business desires because of lack of business context and lack of expertise to connect the dots, incorrect scope and cluster arranged framework. In this paper, we present a goal-oriented big data analytics framework for better business decisions, which consists of a conceptual model which connects a business side and a big data side providing context information around the data, a claim-based evaluation method which enables to focus the most effective solutions, a process on how to use proposed framework and an assistant tool which is a real-time big data analytics platform. In this framework, problems against business goals of the current process and solutions for the future process are explicitly hypothesized in the conceptual model and validated on real big data using big queries or big data analytics. As an exact examination, a shipment choice process is utilized to indicate how IRIS can bolster better business choices as far as extensive understanding both on business and information investigation, high need and quick choices.

Keywords- Big Data Analytics, Goal-Orientated Approach, Business Process, Big Data, Conceptual Model.

I. INTRODUCTION

Big data analytics is a technology which helps turn hidden insights in big data into business value by using advanced analytics techniques in order to support better business decisions, as the hottest new practice in Business Analytics today. According to some surveys, about 80% of CEOs or executive teams view that big data analytics initiatives have the potential to drive business value such as creating new revenue streams, improving operational efficiency or cutting cost, and over 80% of the participant organizations do ongoing projects. However, according to another survey, 55% of big data projects do not get completed, and many others fall short of their business objectives.

Additionally, the quality of big data analytics can only be as good, or as bad, as the quality of the big data it uses. The quality attributes of big data, together with relationships between data, should therefore be defined in a big data model. With a good quality big data model, big data analytics 2
Can accurately identify important business concerns, trend opportunities, and useful business insights, which, in turn, can lead to good business decisions. Yet, no guidelines are available for how to develop a high-quality big data model in a systematic and rational way.

Finally, with the advances in Information Technology (IT), use of software systems increasingly has become prevalent and critical in more efficiently and effectively carrying out the various tasks and activities in the fast changing business domain. However, aligning a software system well with its intended business process has been challenging, due to the difficulty in firstly understanding and precisely modeling a business process and secondly coming up with a requirements specification of a system for supporting the business process. Use of well-known notations helps: BPMN (Business Process Model and Notation) for precise modeling of a business process, and UML use case for modeling software requirements; but the second difficulty still remains. There are several issues: 1) Since there do not exist formal definitions of the models, different interpretations are possible which can lead to transformations that deviate from the original meaning; 2) A business activity can be performed either by people or system functionality; 3) The granularity of a use case is not necessarily the same as that of a business activity or task; and 4) Neither BPMN nor UML use cases consider Non-Functional Requirements (NFRs).

II. RELATED WORK

Literature survey is the most important step in any kind of research. Before start developing we need to study the previous papers of our domain which we are working and on the basis of study we can predict or generate the drawback and start working with the reference of previous papers.

In this section, we briefly review the related work on retailing system and their different techniques.

1] Joseph Amankwah-Amoah,, Samuel Adomako" Big data analytics and business failures in data-Rich environments: An organizing framework" ScienceDirect 2018

The analysis provides insights into how ordinary big data analytic capability and mere possession of big data are more likely to create conditions for business failure. The study extends the existing streams of research by shedding light on decisions and processes in facilitating or hampering firms' ability to harness big data to mitigate the cause of business failures. The analysis led to the categorisation of a number of fruitful avenues for research on data-driven approaches to business failure.

2] Ali Al Hadwer, Dan Gillis, Davar Rezania" Big Data Analytics for Higher Education in The Cloud Era" 2019 the 4th IEEE International Conference on Big Data Analytics.

In this paper, they explore how data and analytics have been used so far in the higher education sector for enhanced learning or to support decisions, what opportunities and challenges surround BDA in this sector

3] S. Supakkul and L. Chung, "Extending Problem Frames to Deal with Stakeholder Problems: An Agent- and Goal-oriented Approach,"Proc., ACM Symp. on Applied Comp. 2009. pp. 389-394.:

The Problem Frames approach captures problems as seen by developers, but not by stakeholders. This paper presents a framework that extends the problem diagram of the Problem Frames approach to represent stakeholder problems using "soft-problem", a notion referring to an undesirable situation that negatively affects stakeholder goals and may have less clear-cut resolution criteria. A soft-problem may be refined to more specific sub-problems and root causes, which are then traced to corresponding solutions in a diagram called Problem Interdependency Graph. The framework has been applied to the 1992 London ambulance case study, which shows that the framework helps to more precisely represent the stakeholder problems suffered by the existing system ("as-is") together with the solutions to be provided by the new system ("to-be") to such problems. The study also shows that the framework helps to determine whether the problems have been sufficiently addressed and, if not, why.

4] Business Process Model and Notation: The Business Process Modeling Notation - BPMN – provides a common language which allows all the parties involved to communicate processes clearly, completely and efficiently. In this way, BPMN defines the notation and semantics of a Business Process Diagram (BPD).

5] A. Davidson and A. Or. "Optimizing shuffle performance in spark." University of California, Berkeley-Department of Electrical Engineering and Computer Sciences, Tech. Rep, 2013:

Spark has been established as an attractive platform for big data analysis, since it manages to hide most of the complexities related to parallelism, fault tolerance and cluster setting from developers. However, this comes at the expense of having over 150 configurable parameters, the impact of which cannot be exhaustively examined due to the exponential amount of their combinations. The default values allow developers to quickly deploy their applications but leave the question as to whether performance can be improved open. In this work, we investigate the impact of the most important tunable Spark parameters with regards to shuffling, compression and serialization on the application performance through extensive experimentation using the Spark-enabled Marenostrum III (MN3) computing infrastructure of the Barcelona Supercomputing Center. The overarching aim is to guide developers on how to proceed to changes to the default values. We build upon our previous work, where we mapped our experience to a trial-and-error iterative improvement methodology for tuning parameters in arbitrary applications based on evidence from a very small number of experimental runs. The main contribution of this work is that we propose an alternative systematic methodology for parameter tuning, which can be easily applied onto any computing infrastructure and is shown to yield comparable if

not better results than the initial one when applied to MN3; observed speedups in our validating test case studies start from 20%. In addition, the new methodology can rely on runs using samples instead of runs on the complete datasets, which render it significantly more practical.

6] J. Horkoff, D. Barone, L. Jiang, E. Yu, D. Amyot, A. Borgida and J. Mylopoulos, "Strategic business modeling: representation and reasoning," Software & Systems Modeling, 2014. pp. 1015-1041:

Business intelligence (BI) offers tremendous potential for business organizations to gain insights into their day-to-day operations, as well as longer term opportunities and threats. However, most of today's BI tools are based on models that are too much data-oriented from the point of view of business decision makers. We propose an enterprise modeling approach to bridge the business-level understanding of the enterprise with its representations in databases and data warehouses. The business intelligence model (BIM) offers concepts familiar to business decision making--such as goals, strategies, processes, situations, influences, and indicators. Unlike many enterprise models which are meant to be used to derive, manage, or align with IT system implementations, BIM aims to help business users organize and make sense of the vast amounts of data about the enterprise and its external environment. In this paper, we present core BIM concepts, focusing especially on reasoning about situations, influences, and indicators. Such reasoning supports strategic analysis of business objectives in light of current enterprise data, allowing analysts to explore scenarios and find alternative strategies. We describe how goal reasoning techniques from conceptual modeling and requirements engineering have been applied to BIM. Techniques are also provided to support reasoning with indicators linked to business metrics, including cases where specifications of indicators are incomplete. Evaluation of the proposed modeling and reasoning framework includes an on-going prototype implementation, as well as case studies.

7] S. Supakkul, L. Zhao and L. Chung, "GOMA:Supporting Big Data Analytics with a Goal-Oriented Approach," IEEE BigData Con.,2016. pp. 149-156:

The real value of Big Data lies in its hidden insights, but the current focus of the Big Data community is on the technologies for mining insights from massive data, rather than the data itself. The biggest challenge facing industries is not how to identify the right data, but instead, it is how to use insights obtained from Big Data to improve the business. To address this challenge, we propose GOMA, a goal-oriented modeling approach to Big Data analytics. Powered by Big Data insights, GOMA uses a goal-oriented approach to capture business goals, reason about business situations, and guide decision-making processes. GOMA provides a systematic approach for integrating two types of the resulting insight from data analytics to goal-oriented reasoning and decision-making processes: descriptive insights are the ones that describe the current state (e.g., the current customer retention rate) and predictive insights are the ones that predict likely future phenomena by inference from the data (e.g., customers who are likely to defect). To aid in the description and illustration of the GOMA approach, a retail banking churning scenario is used as a running example throughout this paper.

8] W. M. P. van der Aalst, Process Mining: Discovery, Conformance and Enhancement of Business Processes, Springer Science & Business Media. 2011:

More and more information about business processes is recorded by information systems in the form of so-called "event logs". Despite the omnipresence of such data, most organizations diagnose problems based on fiction rather than facts. Process mining is an emerging discipline based on process model-driven approaches and data mining. It not only allows organizations to fully benefit from the information stored in their systems, but it can also be used to check the conformance of processes, detect bottlenecks, and predict execution problems. Wil van der Aalst delivers the first book on process mining. It aims to be self-contained while covering the entire process mining spectrum from process discovery to operational support. In Part I, the author provides the basics of business process modelling and data mining necessary to understand the remainder of the book. Part II focuses on process discovery as the most important process mining task. Part III moves beyond discovering the control flow of processes and highlights conformance checking, and organizational and time

perspectives. Part IV guides the reader in successfully applying process mining in practice, including an introduction to the widely used open-source tool ProM. Finally, Part V takes a step back, reflecting on the material presented and the key open challenges. Overall, this book provides a comprehensive overview of the state of the art in process mining. It is intended for business process analysts, business consultants, process managers, graduate students, and BPM researchers.

III. EXISTING APPROACH

A lot of work has been done in this field thanks to its extensive use and applications. This section mentions some of the approaches that have been implemented to achieve the same purpose. These works are mainly differentiated from the techniques for retailing systems.

1) Lack of business context around the data and lack of expertise to connect the dots, which hinder comprehensive understanding of business and analytics results for right decisions.

2) Inaccurate scope with business silo, which means that big data analytics are done regardless of business objectives under rare communications between departments in an organization.

4) Hard to derive actionable business insights or requirements from data.

IV. PROPOSED APPROACH:-

To address the problems, propose a novel big data analytics framework using Spark in a goal-oriented approach for better business decisions, i.e., supporting comprehensive understanding of business and data, high priority and fastness. This framework provides a conceptual model for big data analytics which connects the dots between important concepts of business side such as business goals, problems, solutions, business processes and those of a big data side such as big analytics and big queries to bridge the gap between the two. It helps not only in comprehensive understandings of current and future business, but also communications between stakeholders by helping analyzers explicitly model the concepts. Moreover, in the spirit of goal-orientation, alternatives in problems with the as-is business process and solutions for the to-be business process can be hypothesized in the conceptual model and validated by analyzing big data, and significant ones are selected after trade-off analysis through our evidence-based evaluation method. This process helps analyzers focus on the most effective solutions within the given time and budget. To support our concepts, we implemented IRIS assistant tool which integrates Eclipse Modelling Framework (EMF) for the conceptual model and Apache Spark for a big data analytics which enables real-time processing in a distributed and clustered computing platform, leading to fast business decisions.

System Diagram:

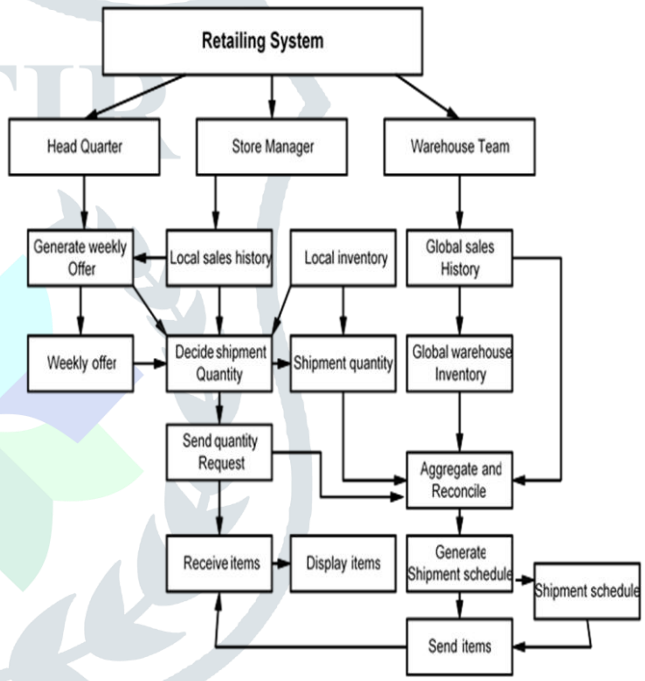

Fig 1. System Architecture

Algorithm:

Naive Bayes

Steps:

1. Given training dataset D which consists of documents belonging to different class say Class A and Class B
2. Calculate the prior probability of class A=number of objects of class A/total number of objects

   Calculate the prior probability of class B=number of objects of class B/total number of objects
3. Find NI, the total no of frequency of each class

   Na=the total no of frequency of class A

   Nb=the total no of frequency of class B

4. Find conditional probability of keyword occurrence given a class:

   P (value 1/Class A) =count/ni (A)

   P (value 1/Class B) =count/ni (B)

   P (value 2/Class A) =count/ni (A)

   P (value 2/Class B) =count/ni (B)

   ……..

   ……..

   ……..

   P (value n/Class B) =count/ni (B)

5. Avoid zero frequency problems by applying uniform distribution
6. Classify Document C based on the probability p(C/W)
   a. Find P (A/W) =P (A)*P (value 1/Class A)* P (value 2/Class A)……. P(value n /Class A)
   b. Find P (B/W) =P (B)*P (value 1/Class B)* P (value 2/Class B)……. P(value n /Class B)
7. Assign document to class that has higher probability.

V. RESULTS AND DISCUSSION

Experiments are done by a personal computer with a configuration: Intel (R) Core (TM) i3-2120 CPU @ 3.30GHz, 8GB memory, Windows 7, MySQL 5.1 backend database and Jdk 1.8, scala and spark. The application is web application used tool for design code in Eclipse and execute on Tomcat server.

A. *Comparison Graph:*

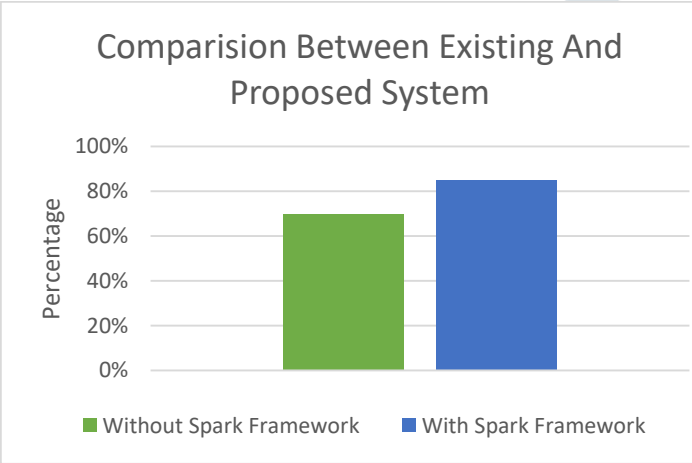

Fig 2. Graph

B. *Comparison Table:*

Sr.No	Existing Method Result	Proposed Method Result
1	65%	87%

Conclusion

We have proposed a goal-oriented big data analytics framework. More specifically, proposed framework includes conceptual model which connects business and big data, an evidence-based evaluation method for selecting the most effective solutions, a process for finding business problems and solutions – firstly by hypothesizing them, and secondly by validating them using big queries or big analytics – and a supporting tool which is implemented on top of Spark, real-time big data analytics framework. Although there are several limitations, at least through an empirical study, proposed can help with big data business analytics in a value-added manner, i.e., comprehensive understanding on business and analytics, high priority, and fast decisions.

REFERENCES

[1] Joseph Amankwah-Amoah,, Samuel Adomako" Big data analytics and business failures in data-Rich environments: An organizing framework" ScienceDirect 2018

[2] Ali Al Hadwer, Dan Gillis, Davar Rezania" Big Data Analytics for Higher Education in The Cloud Era" 2019 the 4th IEEE International Conference on Big Data Analytics.

[3] S. Supakkul and L. Chung, "Extending Problem Frames to Deal with Stakeholder Problems: An Agent- and Goal-oriented Approach,"Proc., ACM Symp. on Applied Comp. 2009. pp. 389-394.

[4] Business Process Model and Notation

[5] Davidson and A. Or. "Optimizing shuffle performance in spark." University of California, Berkeley-Department of Electrical Engineering and Computer Sciences, Tech. Rep, 2013

[6] J. Horkoff, D. Barone, L. Jiang, E. Yu, D. Amyot, A. Borgida and J. Mylopoulos, "Strategic business modeling: representation and reasoning," Software & Systems Modeling, 2014. pp. 1015-1041

[7] S. Supakkul, L. Zhao and L. Chung, "GOMA:Supporting Big Data Analytics with a Goal-Oriented Approach," IEEE BigData Con.,2016. pp. 149-156

[8] W. M. P. van der Aalst, *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, Springer Science & Business Media. 2011