An Internship Report

On

PROCESS MINING VIRTUAL INTERNSHIP

Submitted in partial fulfillment of the requirements for

the award of the degree of

BACHELOR OF TECHNOLOGY

in

Computer Science and Engineering (AI & ML)

by

T. MOIN ALI KHAN (224g1a3357)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
(ARTIFICIAL INTELLIGENCE & MACHINE LEARNING)

SRINIVASA RAMANUJAN INSTITUTE OF TECHNOLOGY
(AUTONOMOUS)
(Affiliated to JNTUA, accredited by NAAC with 'A' Grade, Approved by AICTE, New Delhi & Accredited by NBA (EEE, ECE & CSE))
Rotarypuram village, B K Samudram Mandal, Ananthapuramu-515701.

2024 - 2025

SRINIVASA RAMANUJAN INSTITUTE OF TECHNOLOGY (AUTONOMOUS)

(Affiliated to JNTUA, accredited by NAAC with 'A' Grade, Approved by AICTE, New Delhi & Accredited by NBA (EEE, ECE & CSE))
Rotarypuram village, B K Samudram Mandal, Ananthapuramu-515701.

Department of Computer Science & Engineering (AI & ML)

Certificate

This is to certify that the internship report entitled "Process Mining Virtual Internship" is the bonafide work carried out by **T.MOIN ALI KHAN** bearing Roll Number 224g1a3357 in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science and Engineering (AI & ML)** for 10 weeks from April 2024 to June 2024.

Internship Coordinator

Mr. P. Veera Prakash, M. Tech., (Ph.D.),
Assistant Professor & HOD of CSE

Head of the Department

Dr. P. Chitralingappa, M. Tech., Ph.D.,
Associate Professor

Date:

Place: Ananthapuramu

EXTERNAL EXAMINER

PREFACE

Brief overview of the company's history:

- **Who founded it:** All India Council for Technical Education (AICTE) has initiated various activities for promoting industrial internship at the graduate level in technical institutes and Eduskills is a Non-profit organization which enables industry 4.0 ready digital workforce in India. The vision of the organization is to fill the gap between Academic and Industry by ensuring world class curriculum access to the faculties and students. Formation of the All-India Council for Technical Education (AICTE) in 19445 by the Government of India.
- **What purpose and when:** With a vision to create an industry-ready workforce who will eventually become leaders in emerging technologies, Eduskills & AICTE launches 'Virtual Internship' program on Process Mining. This field is one of the most in-demand, and this internship will serve as a primer.

Company's Mission Statement: The main mission of these initiatives is enhancement of the employability skills of the students passing out from Technical Institutions.

Business Activities: The All India Council for Technical Education (AICTE) primarily focuses on regulating and promoting technical education in India. Its business activities include accrediting institutions, approving new courses, setting quality standards, fostering research, providing policy recommendations, and ensuring the overall development of technical education across the country.

ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without the mention of people who made it possible, whose constant guidance and encouragement crowned our efforts with success. It is a pleasant aspect that I have now the opportunity to express my gratitude for all of them.

It is with immense pleasure that I would like to express my indebted gratitude to my internship coordinator **Mr. P. Veera Prakash, Assistant Professor & HOD, Department of Computer Science and Engineering**, who has supported me a lot and encouraged me in every step of the internship work. I thank him for the stimulating support, constant encouragement and constructive criticism which have made possible to bring out this internship work.

I am very much thankful **to Dr. P. Chitralingappa, Associate Professor & HOD, Computer Science and Engineering (AI & ML),** for his kind support and for providing necessary facilities to carry out the work.

I wish to convey my special thanks to **Dr. G. Balakrishna, Principal of Srinivasa Ramanujan Institute of Technology** for giving the required information in doing my internship. Not to forget, I thank all other faculty and non-teaching staff, and my friends who had directly or indirectly helped and supported me in completing my internship in time.

I also express our sincere thanks to the Management for providing excellent facilities and support.

Finally, I wish to convey my gratitude to my family who fostered all the requirements and facilities that I need.

T.MOIN ALI KHAN

(224g1a3357)

INDEX

List of Figures

List of Abbreviations

PM	Process Mining
RPA	Robotic process automation
CSV	comma-separated values
XML	Extensible Markup Language.
ERP	Enterprise Resource Planning
SRM	Supplier Relationship Management
MRI	Magnetic Resonance Imaging
NPS	Net Promoterter Score
KPI	Key Performance Indicators

CHAPTER - 1

INTRODUCTION TO PROCESS MINING

1.1what is process mining:

Traditional approaches fail to understand the real-life complexity of processes and also struggle to provide complete insights and visibility given the vast amounts of data that are now available. By contrast, Process Mining offers a data-driven and therefore more objective and holistic approach to understanding business processes. As a result, Process Mining has come to dominate a large majority of operational excellence, automation and digitalization ambitions within industry.

Process Mining is the leading new technology when it comes to talking about ALGORITHEMIC BUSINESS – in other words, businesses that use algorithms and large amounts of real-time data to create business value. This has only become possible through the advent of information systems and administrative tools (e.g. Enterprise Resource Planning or Customer Relationship Management systems) which provide a good data source for process analytics.

Only companies that embrace digital transformation and use data insights to optimize their processes will elevate their customer experiences and eventually drive business success. The new generation of insight-driven companies have understood that digital footprints are everywhere and can be used to make customers happier. With the help of these footprints, they can optimize, innovate, and accelerate their products and services as well as the operations at the core of their business. It is therefore not surprising that companies across all kinds of industries are adopting Process Mining.

Fig 1.1: Process Mining

The process mining goal main challenges is to create a consistent and explicit process model given an event log and the use of tools to diagnose issues observing dynamic behavior (van der Aalst & Weijters, 2004). The identification of issues and diagnoses also needs explore the causal and casual (occasional) relations between activities, and this functionality is not present in a traditional Workflow Management System (WFMS) or Business Process Management System (BPMS). BPMS makes event log acquisition easier for process mining applications, but it is also viable to obtain event logs from different electronic transactions, registers, documents, or spreadsheets.

CHAPTER 2

HOW PROCESS MINING WORKS

2.1 How does process mining work?

- Process mining is a powerful approach that leverages data science techniques to unveil, validate, and enhance workflows within organizations. By amalgamating the realms of data mining and process analytics, process mining enables businesses to extract valuable insights from the log data generated by their information systems. This innovative methodology facilitates a comprehensive understanding of process performance, subsequently identifying bottlenecks and areas for optimization, thereby enhancing overall operational efficiency.
- At its core, process mining relies on the systematic analysis of event logs, which are records of various activities performed within an organization's systems. These logs capture intricate details about each step, decision, and interaction that occur during a process, providing a detailed digital footprint of how tasks progress within the system. Traditional process analysis methods might involve manual observation or subjective interviews, but process mining eliminates this subjectivity by employing advanced data science techniques to uncover patterns, trends, and anomalies in the data.
- Process mining's fusion of data mining and process analytics is pivotal in uncovering hidden inefficiencies and opportunities for enhancement. By applying data mining techniques to event logs, organizations can identify patterns and correlations that might not be immediately apparent. For instance, they can detect frequent deviations from the expected process flow or uncover relationships between process variables that influence performance. These discoveries, often unattainable through manual analysis, provide actionable insights for process optimization.

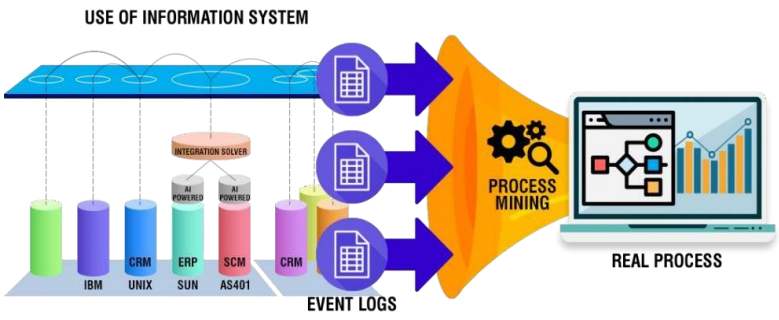

Fig 2.1: How process mining works

- One of the key benefits of process mining is its capacity to offer an objective view of processes as they truly unfold, as opposed to relying on assumptions or perceptions. By processing and analyzing large volumes of event log data, process mining provides organizations with accurate and data-driven insights into their operational reality. This enables them to move beyond relying solely on human understanding and intuition, empowering them to make informed decisions based on empirical evidence.
- Process analytics, on the other hand, focuses on transforming raw event data into meaningful metrics and visualizations. These representations offer a clear and comprehensible view of process performance, enabling stakeholders to identify bottlenecks, delays, and areas where resources are underutilized. Visualizations like process flowcharts, time histograms, and performance heatmaps aid in pinpointing process inefficiencies and irregularities. Armed with this knowledge, organizations can make targeted improvements to streamline processes, allocate resources effectively, and ultimately enhance their overall competitiveness.
- Moreover, process mining is an iterative process, where organizations continuously collect data, analyze processes, and implement improvements. This continuous improvement cycle allows businesses to monitor the impact of changes, adapt to evolving circumstances, and sustain their competitive edge over time.

Nowadays, there are registered events in many business processes, enterprise systems, automation and control systems, medical systems, daily activities, IoT devices, and social networks, among others. These events offer many possibilities to acquire knowledge to understand what is happening *de facto*, or which the most assiduous partners are. This initiative leads to the process mining area that is aimed to discover, check and improve real business processes from events available in many systems (van der Aalst, 2016). Traditional business process design starts from a detailed mapping approach involving multiple resources to establish a consensus model with the most recognized perspective of participants. As an alternative, process mining assumes that it is possible to obtain a meaningful process model extracting it from temporal documents or event logs readily available in system databases.

The process mining goal main challenges is to create a consistent and explicit process model given an event log and the use of tools to diagnose issues observing dynamic behavior (van der Aalst & Weijters, 2004). This functionality is not present in a traditional Workflow Management System (WFMS) or Business Process Management System (BPMS). BPMS makes event log acquisition easier for process mining applications, but it is also viable to obtain event logs from different electronic transactions, registers, documents, or spreadsheets.

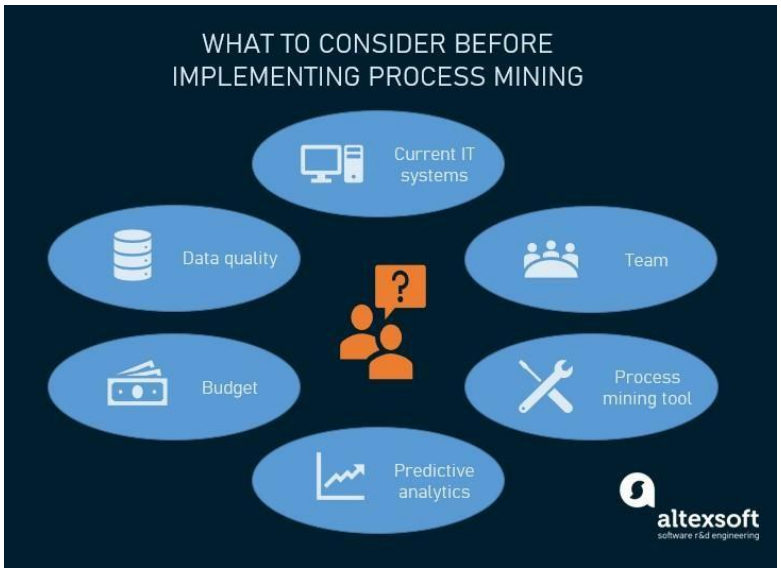

Fig 2.2:implementation of process mining

- The Business Process Management (BPM) approach can be considered an evolution of Workflow Management. BPM is a structured and systematic approach to continuous process analysis.
- According to van der Aalst and Damiani (2015), WFMS emerged in the mid-90s and focused on offering ways to automate some task integrated to a human task and to control the information flow. Some years later, BPMS became an extension of WFMS, which is more focused on operation analysis, management roles, and the work spread in the organization. However, the applications of WFMS or BPMS are very limited in many organizations owing to the difficulties in dealing with semistructured or unstructured processes.
- The concerns on how to identify process optimization and opportunities to achieve better results are continuously increasing; organizations are seeking to reduce time to achieve answers, reduce costs, maximize productivity, balance resource utilization, improve quality, minimize risk, and improve work well being. Business process reengineering (BPR), Kaizen, Value Stream Mapping (VSM), Six Sigma, Lean Thinking.
- Value-Based Management (VBM), and Economic Value Added (EVA) have been adopted to improve efficiency and control business growth (Low, van der Aalst, ter Hofstede, Wynn, & Weerdt, 2017). Even though these approaches contribute to the performance improvement and reduction of costs, they have high failure rates reaching between 60% and 70% in BRP (Park & Kang, 2016). It is a motivation to seek more efficient approaches to process improvement and innovation.

In other side, for discovering patterns in large amounts of data, Data Mining (DM) has emerged. DM is based on techniques and methods focused on processing Big Data. The Big Data technologies, such as Hadoop, Hive, Impala, Spark and Storm, are mostly dedicated to processing large volumes of data for delivering traditional reports or dashboards, focusing specific activities (a slice of a business process) and rewriting the traditional BI (van der Aalst & Damiani, 2015). As Big Datas main research efforts involve

storing and processing, other applications using analytical process identification and analysis of temporal event series patterns are not broadly studied in this area.

On the other hand, there is a need to improve and support business processes in competitive and rapidly changing environments. This manifesto is created by the IEEE Task Force on Process Mining and aims to promote the topic of process mining. Moreover, by defining a set of guiding principles and listing important challenges, this manifesto hopes to serve as a guide for software developers, scientists, consultants, business managers, and end-users. The goal is to increase the maturity of process mining as a new tool to improve the (re)design, control, and support of operational business processes.

Using our road trip analogy again, Process Explorer isn't showing the different routes (variants) that people actually took on a given trip (case). Process Explorer shows us which waypoints (activities) and roads (connections) are the most common along the journey.

CHAPTER 3

RESEARCH METHODS AND ITS TOOLS

Research method

This section covers the applied research method in this work conducted to search, select, extract, classify, and analyze previous work to understand the state-of-art direction of the investigation related to our exploratory research questions.

This study employed systematic mapping to identify, organize, and understand the main contributions of the state-of-art relating to process mining techniques and applications. Systematic mapping studies are designed to provide a wide overview of a research

3.1 Process discovery algorithms:

The section provides a mapping of the algorithms used on the process discovery task. The first initiative to answer this question is categorized by the first published work for each algorithm. Fig. 8 presents an overview of process miners and lists some of the most relevant algorithms, including author and year.

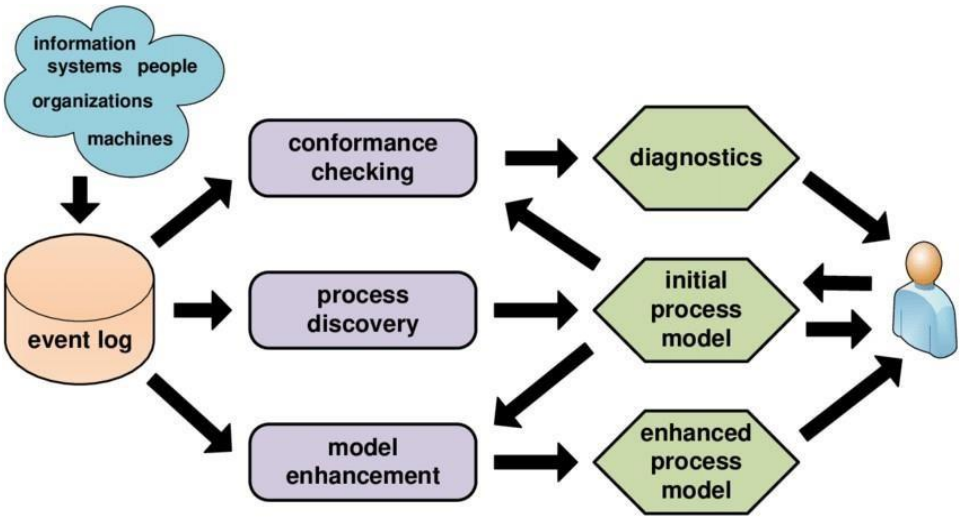

Fig 3.1: discovery of process mining

The first relevant proposal was from Cook and Wolf (1995) describing three methods for process discovery based on finite state machines (FSMs).

This section is aimed at answering the question: Where is process mining being applied? The actual explored process mining application domains are very wide, we could observe that papers exclusively describing some process mining application represents almost 8% of the total number papers. Exploring all the analyzed papers we realized that the most relevant applications are in healthcare industry, hospitals, and clinical path.

Process mining research topics

To answer our research question: "Which research topics can be identified in the primary studies of process mining?" all selected studies were analyzed and categorized according to highest contribution, as some papers included a secondary focus, we also identified a secondary categorization to the papers that have relevant contributions in two research topics, here also called categories. The systematic mapping explored the three main types (van der Aalst et al., 2012) of process mining:

Discussion:

This systematic mapping study established research questions formulated in Section 2. With respect to RQ1: Which research topics can be identified in the primary studies of process mining? we started mapping the main types of process mining: discovery - focused on producing a real process model based on event logs; conformance - comparing a process model (*a priori* or discovered) and an event log, and vice versa; enhancement - to extend results with emphasis on frequencies, working time, waiting

Conclusion for research methods:

The research method applied in this work delivered a breadth-first review of the primary research studies in process mining and provided an overview of the published applications domain. Evaluation of 3713 published papers over the last 16 years resulted in the selection of more than 34% of the papers related to process mining. This large base allowed us to produce a comprehensive map regarding the established research questions.

The Best Process Mining Tools

APPIAN:

Description: Appian is a leading low-code platform provider that allows both experienced and citizen developers to build process-centric and case-centric applications with the ability to monitor and improve business processes in response to changing needs. With Appian's process mining capabilities, organizations can integrate data from multiple systems, identify process bottlenecks, develop purpose-built dashboards for specific analysis needs, predict process behaviors, design optimized workflows, maintain compliance with process standards, reduce operational costs, and more.

BIZAGI:

Description: Bizagi is a leader in digital business process automation software. The vendor offers three tiers of solutions, including Bizagi Engine, Bizagi Studio, and Bizagi Modeler. Bizago's process mining capabilities are included via the Enterprise model of Bizagi Modeler, which equips companies with the process mining tools they need to understand their processes. Other capabilities available with Bizagi Modeler Enterprise include valuechain diagrams, Single Sign-On, model sharing, private cloud storage, real-time notifications, and more.

BONITASOFT:

Description: Bonitasoft develops BPM software for developers to build business applications that adapt to real-time changes, UI updates, and more. With Bonitasoft, users can automate, model, and monitor business processes to streamline operations. The software automatically checks for errors and highlights them before users save their business model. Companies can use Bonitasoft's AI-powered process mining algorithms to analyze data, improve visibility, identify patterns, track performance indicators, define business operating models, predict issues, and create opportunities for improvement.

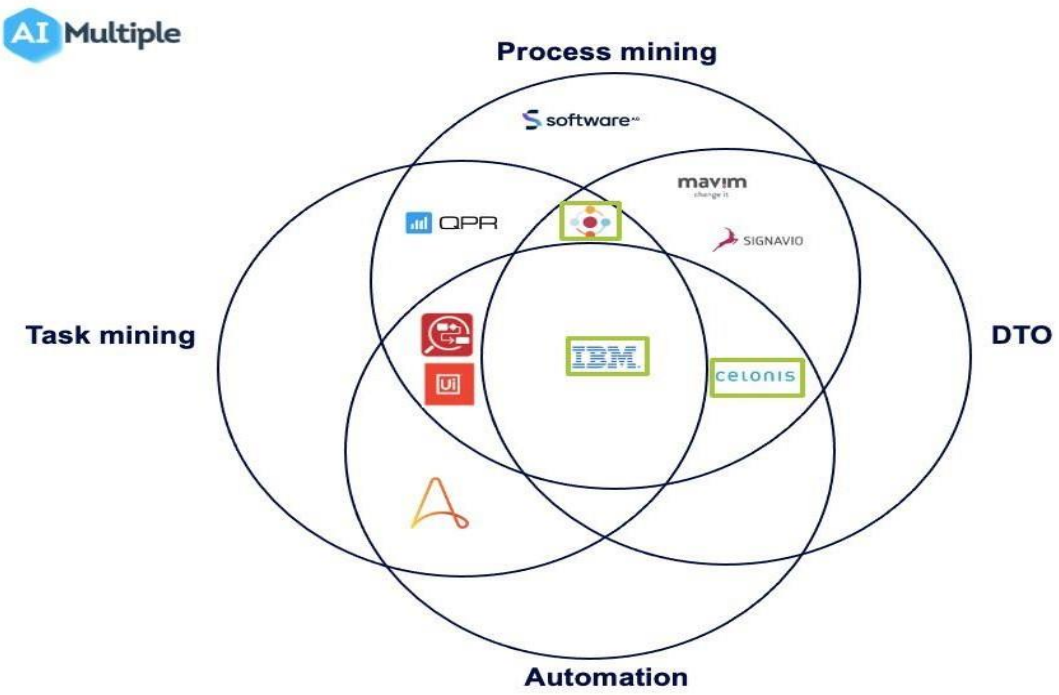

Fig. 3.2: Process mining tools

Celonis:

Description: Celonis is a global provider of execution management solutions that help companies improve how they run their business processes. With Celonis' suite of process and task mining capabilities, companies across industries can improve visibility into their operations, identify bottlenecks, and streamline efficiencies. Those capabilities—powered by machine learning and industry-standard process query language (PQL)—include analytic visualizations, drag-and-drop customization tools, task mining, extensible data models, multi-event logs, best-practice benchmarking, and tools for identifying processes that could benefit from automation.

FLUXICON:

Description: Fluxicon is a process mining solution provider for business process managers and consultants. The company's process mining product, Disco, can help users reduce costs, improve quality, compare processes beyond KPIs, and create high-level models of their processes. Its capabilities include process map animations, detailed statistics, interactive charts, automated process discovery, user-friendly log filters for drilling deeper into data, project management, performance filters, and multiple options for importing and exporting data.

IBM:

The IBM Process Mining product suite uses data-driven process insights to help companies across markets improve processes and make faster, more informed decisions. IBM's process mining tools can be applied in use cases like intelligent automation, customer onboarding, procure-to-pay (P2P), accounts payable, IT incident management, and orderto-cash. Features include automated robotic process automation (RPA) generation, factbased process models, AI-powered process simulations, conformance checking, task mining, and seamless integrations with leading software SAP, Oracle, and other IBM products.

CHAPTER 4

MODULES EXPLANATION

1. Data Extraction and Collection:

- In this initial phase, relevant data is extracted from various sources, such as ERP systems, databases, and application logs.
- The data collected usually includes timestamps, activity names, case IDs, and other relevant attributes that capture the process execution.

2. Data Preprocessing:

- Raw data often requires cleaning and preprocessing to ensure accuracy and consistency.

Data preprocessing involves tasks like handling missing values, removing duplicates, and transforming data into a suitable format for analysis.

3. Event Log Preparation:

- The preprocessed data is structured into event logs, which contain chronological records of activities associated with case IDs.
- Each event in the log consists of a timestamp, activity name, and case ID, providing a comprehensive overview of process execution.

4. Process Discovery:

- Process discovery aims to create a visual representation of the process based on the event logs.
- Various algorithms are used to generate process models, such as Petri nets, process trees, and flowcharts, reflecting the actual flow of activities.

5. Conformance Checking:

- Conformance analysis compares the actual process execution recorded in the event logs with the expected process model.
- Deviations, bottlenecks, and variations are identified to pinpoint areas where the actual process diverges from the intended model.

6. Performance Analysis:

- This module focuses on evaluating process performance metrics like cycle time, throughput, and resource utilization.
- By analyzing these metrics, inefficiencies and areas for improvement can be identified.

7. Enhancement and Optimization:

- Using insights gained from the previous modules, organizations can optimize their processes.
- Process optimization involves redesigning workflows, reallocating resources, and improving overall efficiency based on data-driven recommendations.

8. Visualization and Reporting:

- Effective visualization tools are used to represent process models, performance metrics, and conformance analysis results.
- Clear visualizations help stakeholders understand complex process dynamics and make informed decisions.

9. Predictive Analysis:

- Some advanced process mining tools incorporate machine learning techniques to predict future process behaviors based on historical data.
- Predictive analytics can be used to forecast potential bottlenecks or delays and make proactive adjustments.

10. Continuous Monitoring:

- After implementing process improvements, continuous monitoring ensures that the optimized processes are maintained over time.
- Ongoing analysis helps identify new issues and opportunities for further enhancements.

CHAPTER 5
APPLICATIONS OF PROCESS MINING

General Processes

Process discovery for automation:

Automation provides faster and lower-cost solutions. However, companies need to examine their business processes to use automation tools, such as robotic process automation (RPA) efficiently. Process mining vendors claim that their technology can reduce automation implementation time by 50%.

Organizational Mining:

Process logs can identify organizational relationships, performance gaps, and best practices. However, almost all processes have a human component. Process data can be used to understand and improve the human aspects of business processes.

Sales

Reduced sales cycle time:

Lead-to-order processes can take a long time. This causes the payback time of marketing investments to increase. Companies can uncover the reasons behind this issue and, take action to reduce sales cycle time.

Increased conversion rate:

Converting marketing strategies into sales is critical for companies. With a process mining tool, companies can discover if they have proper strategies for increasing conversion rates.

IT Service Management

Reduced risk in ERP related developments:

In Lassila & Tikanoja's case study, the company has implemented a new ERP system by employing process mining. The company achieved their goal to reduce the risks by increasing visibility to the ERP system and operational processes.

Reduced costs in ERP maintenance, development and support:

Process mining can pinpoint mistakes or gaps in the IT systems, such as SAP. The same case study (Lassila & Tikanoja) showed that the company reduced their implementation costs along with the risks of ERP deployments even though it was not primary goal of the project.

Delivering higher first-time resolution:

IT systems may not provide the correct solution at their first try. Process mining tools can produce data-driven insights to increase the first-time resolution rate.

CHAPTER 6

LEARNING OUTCOMES

After completing this Training Track, you will be able to:

- Interpret process visualizations and leverage analyses to identify process inefficiencies.
- Conceptualize your process in terms of activities and cases.
- Save an analysis selection for future reference and share it with your team; export visualizations and process data.
- Perform the basic tasks necessary to build Celonis analyses.
- Become familiar with Analysis Settings and Permissions.
- Publish analyses using best practices in version control.
- Put your knowledge about the theoretical foundations of Process Mining into practice.

CONCLUSION

In conclusion, process mining is a powerful and versatile technology that offers valuable insights into the inner workings of organizational processes. By analyzing event data generated during the execution of processes, process mining uncovers hidden patterns, identifies inefficiencies, and provides actionable recommendations for process optimization. This technology has the potential to drive improvements across a wide range of industries, including manufacturing, healthcare, finance, logistics, customer service, and more.

Process mining's ability to visualize process flows, detect bottlenecks, and pinpoint deviations from the ideal path enables organizations to make informed decisions aimed at enhancing efficiency, reducing costs, and improving overall performance. The real-time applications of process mining are particularly noteworthy, as they empower businesses to respond promptly to changing circumstances, address issues as they arise, and ensure that processes operate at their optimal levels.

As technology continues to advance, process mining techniques are likely to become even more sophisticated and integrated with other data-driven approaches, further enhancing their ability to drive process excellence. However, successful implementation of process mining requires a comprehensive understanding of both the technology and the underlying business processes. Organizations that embrace process mining stand to gain a competitive edge by harnessing the power of data-driven insights to continuously refine their operations and achieve higher levels of efficiency and effectiveness.

INTERNSHIP CERTIFICATE:

REFERENCES:

- Reference Website : https://academy.celonis.com/learningpaths/execution-management.
- Execution management link : https://tinyurl.com/2pvuuzmd.
- Login page : https://academy-login.celonis.com/s/login/.