An Internship Report

on

Process Mining Virtual Internship

Submitted in partial fulfilment of the requirements

for the award of the degree of

BACHELOR OF TECHNOLOGY

in

Computer Science and Engineering (Data Science)

by

K.VIGNA PRIYA RUKMINI   (224G1A32B5)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
(DATA SCIENCE)

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, BKSamudram 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, BKSamudram Mandal, Ananthapuramu-515701.

Department of Computer Science & Engineering (Data Science)

Certificate

This is to certify that the internship report entitled "Process Mining Virtual Internship" is the bonafide work carried out by **K.VIGNA PRIYA RUKMINI** bearing Roll Number 224G1A32B5 in partial fulfilment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science and Engineering (Data Science)** for four months 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

Celonis is actually a German software company, but it is strongly internationalised, with headquarters in Munich and New York.Founded in 1811 by three university students, Celonis market leading platform (Gartner, 1819) has helped companies achieve process excellence by removing operational friction points with its Intelligent Business Cloud platform.

Celonis has developed software that does process mining. Process mining is a technology that allows processes to be analyzed based on data. Processes such as deliveries, productions, or even orders can be recorded and evaluated. The aim behind this is to uncover inefficiencies and uncover potential areas for improvement. Especially in huge companies with cross-departmental processes, it is almost impossible to know, let alone understand, the entire process at a certain point.

Celonis software is called an "Execution Management System". In concrete terms, this means that digital processes are automatically recorded and mapped. This is to create a completely objective and data-driven picture of the company's internal processes. The data for this is usually already available in software such as SAP.

Celonis offers a comprehensive suite of tools that enable businesses to: Discover processes, Diagnose Issues, Optimize Performance, Monitor in Real-Time, Predict and Simulate.

"Process mining delivers insights that enable faster, smarter decisions and stronger performance on an organization's most critical priorities."

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 (Data Science),** 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.

K. Vigna Priya Rukmini
(224G1A32B5)

INDEX

Contents	Page No

LIST OF FIGURES

LIST OF ABBREVIATIONS

AICTE	All India Council for Technical Education
KPI	Key Performance Indicators
EMS	Execution Management System
PQL	Process Query Language
PU	Pull-Up Functions

CHAPTER - 1

INTRODUCTION

Process Mining is the combination of two disciplines: Data Science and Business Process Management. Process Mining essentially uses Data Science techniques, such as Big Data and AI, to address Process Science problems such as process improvement and automation. Process Mining is the leading new technology when it comes to talking about algorithmic businesses - 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.

According to MarketWatch(opens in a new tab), Global Process Mining Software Market is valued approximately at USD 322.02 Million in 1818 and is anticipated to grow more than 50.1% by 1827. The strong development of Celonis - from a student start up to a company with over 3000 employees in 1822 and a customer base of the biggest enterprises like Coca-Cola, Unilever, Vodafone or Uber - paints the same picture. Process Mining is in high demand, which is further backed up by current hypes around automation and other performance acceleration measures (cf. Gartner 1818).

Process Mining is achieved by taking the digital footprints that are created in IT systems and using them to reconstruct and visualize process flows. From here, Process Mining technology can identify patterns and deviations and ultimately eliminate bottlenecks.

CHAPTER-2

PROCESS MINING FUNDAMENTALS

Process mining is an analytical discipline for discovering, monitoring, and improving processes as they actually are and not as you think they might be. Process Mining works by extracting knowledge from event logs (also called digital footprints) readily available in today's information systems, in order to visualize business processes—and their every variation—as they run.

The Celonis Execution Management System (EMS) extends process mining by executing on insights automatically and orchestrating your existing technologies. The Review and Interpret Analyses (opens in a new tab) training track is designed for data and business analysts, process experts, and process improvement specialists. Process Mining helps to build the analyses by using the available tools like variant explorer, process explorer, charts and tables, selection views. It provides Analysis Sheets where the user can extract useful data.

The basic terms in this course includes:

➢ **Process** is a series of linked steps taken in order to achieve a particular goal.

➢ **Activity** is a step that occurs in the process. Process activities are actions that initiate or terminate a process or take place during it. Each activity consists of one or more tasks that together are a milestone in the process.

➢ **Case** is an "item" or "object" you follow through the process. Even for the same business process, the case differs from company to company, depending on how granular they want to get.
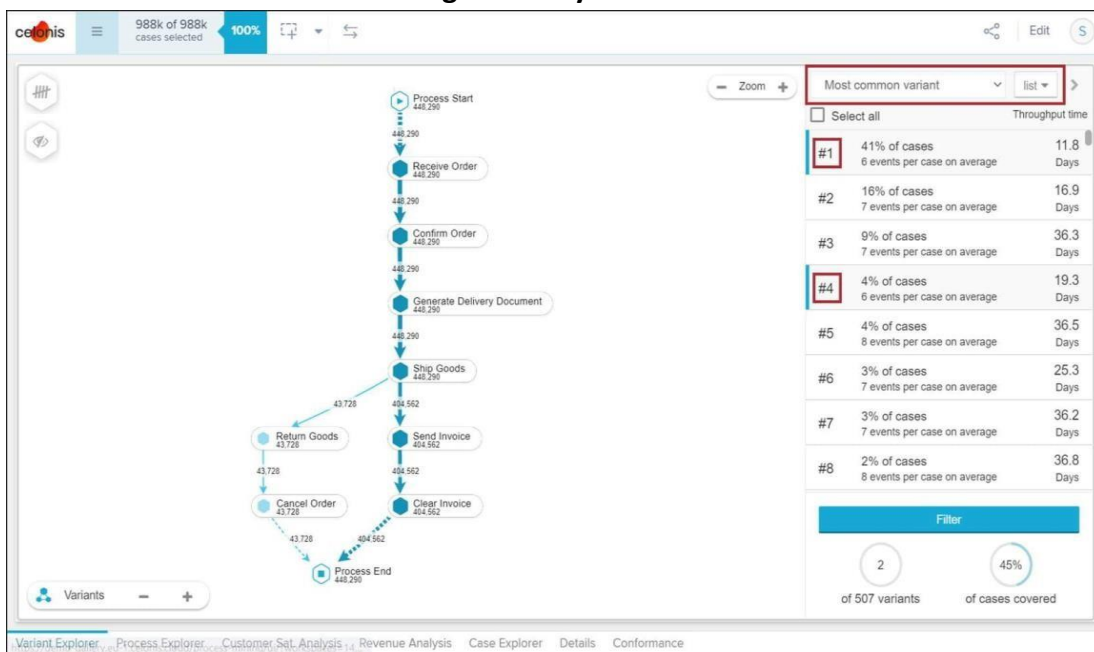
CHAPTER-3

REVIEW & INTERPRET ANALYSIS - I

3.1 Variant Explorer

Using the Variant Explorer, you can discover all the process variants—that is all the different ways the process flows in your organization. The Variant Explorer is one of the Analysis tools to help you take an "exploratory" approach to find out how your process is performing.

Fig 3.1 Analysis Sheet

3.2 Process Explorer

The Process Explorer is another analysis tool to use when taking an exploratory approach. It's especially useful for quickly revealing activities beyond the most common ones. It also allows you to narrow your focus on a single activity, for example an undesired activity, to see which activities cases typically come from and which activities they're going to.

Clicking on an activity in the Process Explorer, you can see a list of the predecessor and successor activities. Process Explorer is the List view of activities and connections. This way, you can quickly scroll through a list of the most common activities and connections. And from there, you can even reveal the predecessor and successor activities.

3.3 Charts,Tables & KPIs

- **A dimension** is a category of attributes; for example, the dimension "customer name" is a category for individual customer names. Other examples of dimensions, depending on the nature of the process, can include vendor name, sales organization, region, and material group.

- **Key Performance Indicators (KPIs)** are used to calculate and add aggregated values; for example, case count, order value, invoice value, throughput time, and automation rate.

  KPIs may also appear as standalone numbers as seen below with the On-Time Delivery and Net Promoter Score (NPS) examples; we call these analysis components "Single KPIs."

CHAPTER - 4

REVIEW & INTERPRET ANALYSES – II

4.1 Selection Views

Selection Views offer a more comprehensive set of options to filter on cases as compared to what you can do in the components in analysis sheets.

You can access the six Selection Views from anywhere in the analysis by clicking on the Selection Views button located in the analysis toolbar.

Fig 4.1 Selection Views
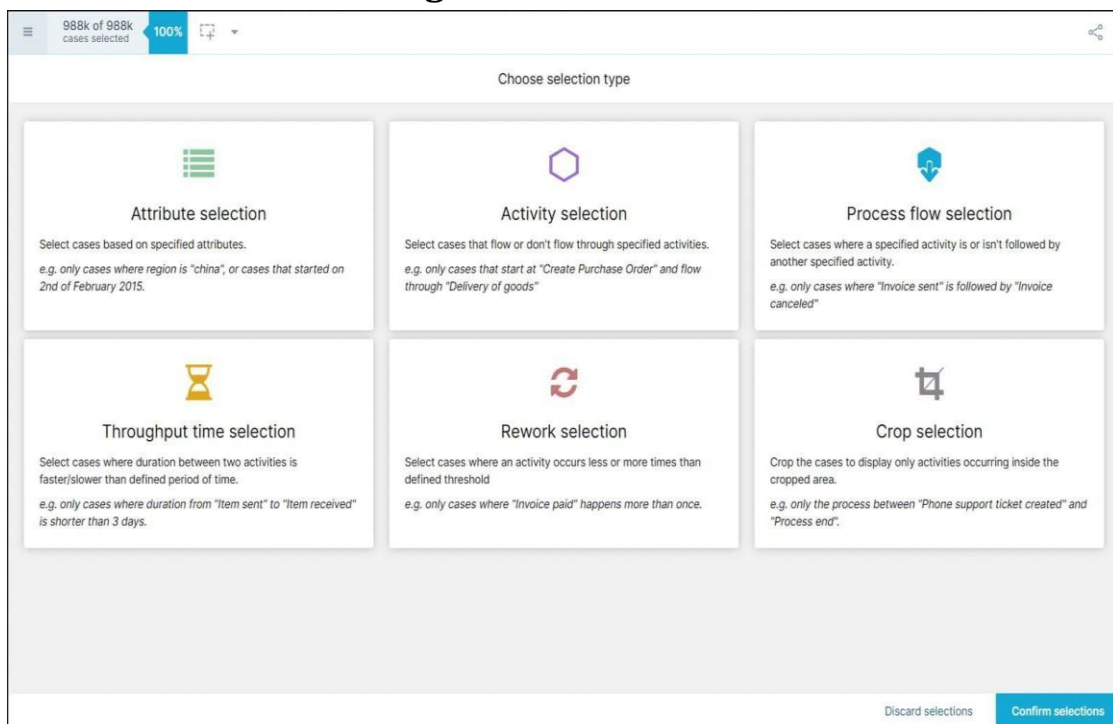
4.2 Case Explorer

The Case Explorer is useful once you've narrowed down the analysis to a few cases that you want to investigate further. You can view specific case details such as timestamp of activities, user type (manual or automatic), possibly even user name (depending on your setup), and other useful pieces of info.

4.3 Conformance Checker

You can use the Conformance checker to:

- Get perspective on the percentage of cases that conform to the idea flow of activities specified in the Analysis (target process model).

- get automated insights into potential root causes of inefficiency by reviewing the most common process violations and the attributes they're associated with.

- mark certain process violations as acceptable to include them in the conforming cases statistics.

The Conformance checker evaluates each case against the process model your organization has specified to determine whether it conforms to it or not.

Conformance Check takes into account two types of criteria in the Violations list:

- Activities not reflected in the process model

- Order of activities not reflected in the process model

4.4 Save & Share Analysis Selection

In Celonis Analysis, you can export data and even the process visualization, if enabled by the person building the analysis.

Right-click on the component to see your options as they differ depending on the component.

CHAPTER – 5
BUILD ANALYSIS
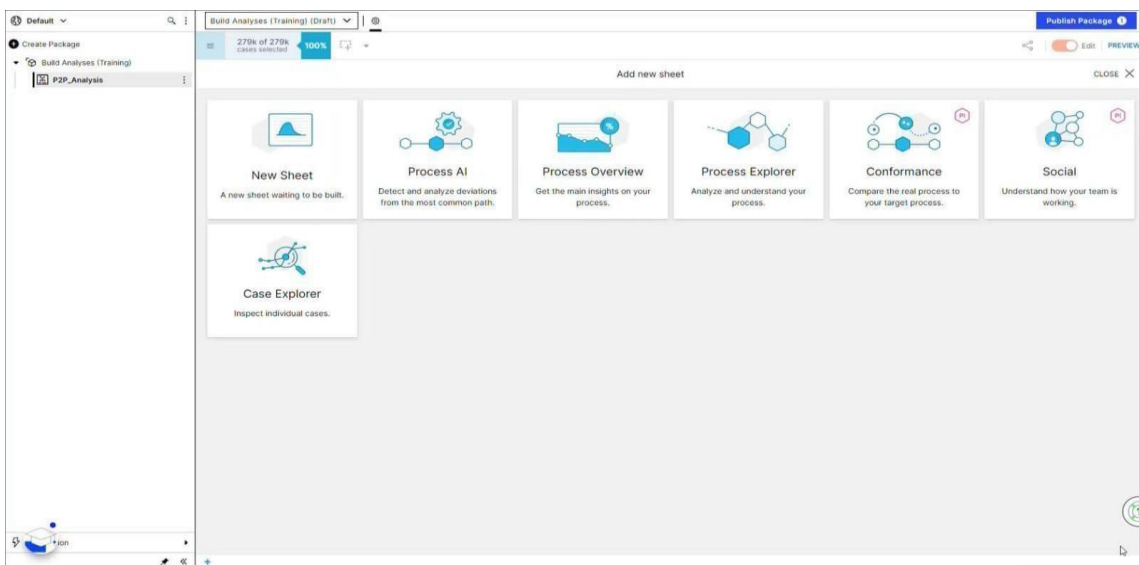
5.1 Analysis Sheet

You'll build analyses in Studio and users will view the published analyses in Apps.

Fig 5.1 Analysis Sheet options

5.2 Configure Tables, Charts &KPIs

Three data tables we used to configure

- OLAP Tables
- Column and Line Charts
- Pie Charts

Anytime you add a table or chart to the analysis, you'll need to select the Dimension(s) and KPI(s) to display

➢ Celonis Analysis includes four types of single KPI components. The most common use cases for the single KPI component include the case count and net value.

➢ Aside from the Number, you might choose other Single KPI components such as Gauge, Fill, and Radial, depending on what you need to display.

5.3 Background Filters

Background filters can be applied at three levels:

- A component (such as a Process Explorer or an OLAP Table)

- A sheet

- The entire Analysis

CHAPTER-6
BASIC QUERIES IN PQL

6.1 Celonis PQL Engine

Celonis PQL is an integral component of the Celonis Software Architecture. All Celonis applications use this language to query data from a data model.

Celonis PQL follows four language features:

- First, operators usually create and return a **single column** that is either added to an existing table (e.g., the case or activity table) or to a new, temporary result table.

- Second, the **supported data types** comprise STRING, INT, FLOAT, and DATE.

- Third, **Boolean values** are not directly supported, but can be represented as integers.

- Fourth, each data type can hold **NULL values**. Celonis PQL operates

Fig 6.1 Celonis PQL Engine
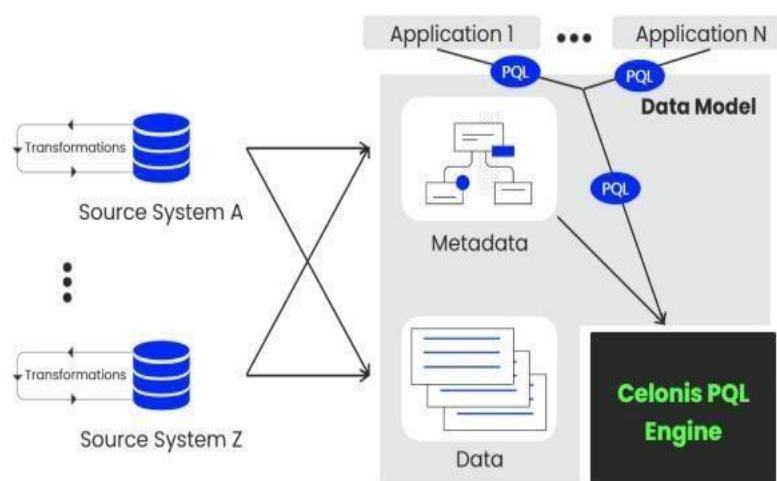
6.2 Basic PQL   Queries

The Data Explorer is an excellent tool that allows you to easily validate your data and build your first PQL queries.

Standard Aggregation Functions

- Counting COUNT()
- Evaluating includes AVG(),MIN(),MAX(),SUM(),MEDIAN()
- QUANTILE(),VAR(),STDEV()

Data Flow Functions
- CASE WHEN
- COALESCE
- REMAP_VALUES

String modifications comprise a set of powerful functions including combinations, transformations and cropping to get the most out of your Celonis queries.

- LOWER,UPPER,REVERSE
- LEFT,RIGHT,LTRIM,RTRIM,SUBSTRING

Data Type conversions
Different types (e.g. integer, float, string or date).

RegEx or Regular Expressions are a way to search within a text according to a particular search pattern. Regular Expressions are built from a sequence of characters that define this search pattern.

CHAPTER-7
JOINING & AGGREGATING DATA

7.1 Joins & Filters

The set of tables containing all the data is called the **Data Model**.

The tables in a Data Model are connected via specific relationships to associate rows of one table with rows of another table. This is done using a **foreign key**. In general, these relationships can be classified as:

- One-to-many or 1:N
- One-to-one or 1:1
- Many-to-many or N:M

A PQL query is executed in the following order:

1. Joins and regular PQL functions (not aggregations). The common table is defined after joining the required tables.
2. Filters are applied (if there are filters defined). We will learn more about filters in chapter 4.
3. Standard Aggregations (AVG, COUNT, SUM, etc.).

Filters:

Center the attention on a given subset of data in order to better understand the information or view it from different angles.

The syntax is very simple. You start with the keyword FILTER, followed by the condition. An easy condition starts with the table and column you'd like to apply the filter to, followed by the predicate operator, which could be BETWEEN, !=, <, >=, <=, =, or many others. To complete the condition, you add the respective value that you would like to filter on. Finally, the FILTER statement is terminated with a semicolon.

7.2 PU_ Functions

PU-functions can be applied in many use cases and are often the only way to calculate a certain KPI or dimension.

- PU_SUM
- PU_AVG
- PU_COUNT, PU_COUNT_DISTINCT
- PU_MAX, PU_MIN
- PU_MEDIAN
- PU_QUANTILE
- PU_FIRST, PU_LAST
- PU_STRING_AGG

o PU-functions are key when working across multiple tables that are connected with a 1:N relationship.

o PU-functions over standard aggregations is that PU-functions can be used inside filters, whereas standard aggregations can't be used in this context

o PU-functions are one of the best possibilities for solving the "no common table error".

CHAPTER - 8
SET UP A DATA PIPELINE

8.1 Data Pipeline

It helps you connect to source systems, extract the relevant data, transform it to your needs, and load it into a polished Data Model.

- validation, scheduling, monitoring, optimizing performance Celonis offers real-time extraction and transformation capabilities. In simple terms, real-time means the EMS can track and frequently replicate incremental changes in data from source systems .

The main objective behind real-time pipelines is for users to operationally act on the data.

8.2 Data Integration in the EMS

Data Integration is where you set up connections and your data pipeline.

The main ways you can bring data into the EMS are:

1. Process Connectors

   Process Connectors are the fastest and most common way for you to connect and cover the most established process use cases.

2. Extractors (Data Connections)

   Extractors are blank data connections with no reference to a process. With Extractors, you connect to source systems and then have to build your data pipeline from scratch—i.e. your extractions, transformations, and Data Model.

3. Extractor Builder

This tool helps you quickly build an Extractor using REST API and supports the OData V2 and V4 standards as well. It allows you to quickly connect to source systems and extract the data you need.

4. File Uploads

- For additional static data
- a list of translation terms
- For a list of permissions outside of your source systems
- For historical data that you know will never change
- For one-time uploads

5. Data Push API

The Data Push API uses micro batching (data chunks) to provide a scalable and reliable integration point and allows Parquet and CSV file formations. The Data Push API is a set of sequential API calls to create, execute and monitor "Push Jobs"

6. Celoxtractor

The Celoxtractor is a Python package designed to let you develop your own EMS Extractor easily. It gives you:

- complete control over your data,
- feature parity to native Celonis extractors,
- and full flexibility in adjusting all aspects of your extractions.

CHAPTER - 9
REFINE DATA PIPELINE

9.1 Schedule Data Jobs

- Extractions and Transformations and Data Model Loads in a **continuous manner**.

- For a continuous and automated data load you can use the scheduling functionality. Schedules allow you to sequentially execute Data Jobs on a regular basis.

9.2 Troubleshoot & Monitor

- It's important for you to know the ins and outs of troubleshooting your Data Pipeline from the moment you connect to source systems to when you set up your Data job schedules or replications using the Replication Cockpit.

- If your Data Pipeline is set up end to end, then it is theoretically time for you to sit back and enjoy the beauty of your work. That said, you should make sure to keep an eye on things to ensure everything keeps running smoothly.

9.3 Extract Necessary Data

An optimized data pipeline extracts and transforms **only necessary data.** Skipping an optimization of your extractions can have very detrimental effects on your data pipeline. It leads to:

- Higher storage needs (affects your license's APC)
- Heavier consequences of bad practices in transformations (e.g. SELECT *)

- Negative performance impact on load times in extractions, transformations, and data model loads

For each query submitted to Vertica, the Vertica query optimizer assembles **a query execution plan**—a sequence of steps and required operations to access data and calculate the result.

9.4   Custom Processes

- A unique identifier makes it possible to distinguish between different cases and correctly assign the activities and timestamps. In the Event Log, we call such a unique identifier a "**Case ID**."
- Single process can be **parallelly executed** in multiple systems or **sequentially executed** in multiple systems.

Four important points to consider when defining data requirements:

- **Activity Data** - Used as the main ingredient for process mining to generate an event log.
- **Dimensions** - Show the process/metrics for specific attributes such as vendors and product categories.
- **Key Metrics** - Allow us to align on the most important calculations prior to data extraction.
- **Translation & Name Mappings** - Convert certain technical terms into meaningful text fields.

Finally ,run a **quality assurance check on your Data Pipeline** before it goes live. The checklist is based on project best practices and currently used in implementations.

CONCLUSION

Process Mining is achieved by taking the digital footprints that are created in IT systems and using them to reconstruct and visualize process flows. From here, Process Mining technology can identify patterns and deviations and ultimately eliminate bottlenecks

Process mining techniques have been used to improve process flows across a wide variety of industries.

- **Education:** Process mining can help identify effective course curriculums by monitoring and evaluating student performance and behaviours, such as how much time a student spends viewing class materials.

- **Finance:** Financial institutions have used process mining software to improve interorganizational processes, audit accounts, increase income, and broaden its customer base.

- **Public works:** Process mining has been used to streamline the invoice process for public works projects, which involve various stakeholders, such as construction companies, cleaning businesses, and environmental bureaus.

- **Software Development:** Since engineering processes are typically disorganized, process mining can help to identify a clearly documented process. It can also help IT administrators monitor the process, allowing them to verify that the system is running as expected.

N·E·A·T
प्रौद्योगिकी के लिए राष्ट्रीय शैक्षणिक सहयोग
National Educational Alliance for Technology

AICTE

अखिल भारतीय तकनीकी शिक्षा परिषद्
All India Council for Technical Education

EduSkills®
Nation Building Through Skills

AICTE - EduSkills
2024
VIRTUAL INTERNSHIP

Certificate of Virtual Internship

This is to certify that

Kethireddy RUKMINI

Srinivasa Ramanujan Institute of Technology

has successfully completed 10 weeks

Process Mining Virtual Internship

During April - June 2024

Supported By **celonis**

A. Gebert

Angela-Sophia Gebert
Global Head of Academic Alliance
Celonis

Shri Buddha Chandrasekhar
Chief Coordinating Officer (CCO)
NEAT Cell, AICTE

Dr. Satya Ranjan Biswal
Chief Technology Officer (CTO)
EduSkills

Certificate ID :9cd9154964359c0ec28f24fc382f560d
Student ID :STU65897f04b8f461703509764

INTERNSHIP GRADE
B

GRADE- O (Outstanding):90-100 | E (Excellent):80-89 | A (Very Good):70-79 | B (Good): 60-69 | C (Fair): 50-59 | D (Average): 40-49 | P (Pass): 30-39 | F (Fail): Below 30

REFERENCES

[1] https://academy.celonis.com/learn/dashboard

[2] https://academy.celonis.com/pages/academic-rising-star-technical

[3] https://academy.celonis.com/learn/learning-path/process-mining-fundamentals-forstudents

[4] https://academy.celonis.com/learn/learning-path/get-data-into-the-ems-