Dissertation Title:

INTELLIGENT ROLE MINING AND OPTIMIZATION IN IDENTITY AND ACCESS MANAGEMENT

**Course No. : SS ZG628T**

**Course Title: Dissertation**

**Dissertation Done by:**

**Student Name: GODLAVEETI ANIL GOVIND**

**BITS ID: 2023MT12272**

**Degree Program:M.Tech.(Software Systems)**

**Research Area: Machine Learning**

**Dissertation / Project Work carried out at:**

**Bridgesoft Solutions, Hyderabad.**



**BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI**

## VIDYA VIHAR, PILANI, RAJASTHAN - 333031.

**MARCH 2025**

**Contents**

[1. Broad Area of Work 3](#_heading=h.30j0zll)

[2. Background 3](#_heading=h.1fob9te)

[3. Objectives 3](#_heading=h.3znysh7)

[4. Scope of Work 4](#_heading=h.2et92p0)

5. Methodology 4

[6. Plan of Work 1](#_heading=h.tyjcwt)1

[7. Literature References 1](#_heading=h.3dy6vkm)2

[8. Particulars of the Supervisor and Examiner 1](#_heading=h.1t3h5sf)3

[9. Remarks of the Supervisor 1](#_heading=h.4d34og8)4

# Broad Area of Work

Identity and Access Management (IAM) plays a crucial role in ensuring secure access to enterprise systems. The project focuses on enhancing IAM functionalities using SailPoint and advanced techniques in Artificial Intelligence, Data Analytics, and Security Optimization. The key application areas are:

* + - Role Mining in SailPoint: To discover and optimize roles based on user access patterns in SailPoint IdentityIQ (IIQ).
    - Anomaly Detection: To identify unauthorized access or irregular patterns in user behavior.
    - Machine Learning: To develop models that improve SailPoint’s role management and decision-making capabilities.
    - Visualization Tools: To represent data insights effectively for governance and audit purposes within SailPoint.

# Background

Managing user identities and access privileges in large organizations is complex, especially as digital infrastructures grow. SailPoint IdentityIQ (IIQ) is a leading IAM solution that helps organizations streamline access management, ensure compliance, and secure critical assets.

However, traditional role management in IAM systems, including SailPoint, often relies on manual processes that can lead to inefficiencies and errors. By integrating machine learning into SailPoint’s role mining and optimization framework, organizations can automate role discovery, detect anomalies, and enhance the security and efficiency of their IAM systems.

# Objectives

The primary objectives of this project are:

* + - To analyze access data within SailPoint IdentityIQ to identify access patterns and optimize role definitions.
    - To integrate machine learning models for automated role mining in SailPoint.
    - To develop an anomaly detection mechanism to enhance SailPoint’s ability to detect unauthorized access.
    - To build a visualization dashboard for SailPoint administrators to simplify access management and governance.
    - To validate the solution through testing and integration with the SailPoint platform.

# Scope of Work

Scope of this dissertation includes designing and implementing a machine learning-powered solution tailored to **SailPoint IdentityIQ**. The deliverables are:

1.**Role Discovery in SailPoint**: Utilizing clustering techniques to identify redundant and efficient roles within SailPoint.

2.**Anomaly Detection**: Enhancing SailPoint’s capabilities to detect deviations in access behavior using machine learning.

3.**Dashboard Development**: Designing an intuitive visualization tool for better role management within SailPoint.

4**.System Integration**: Ensuring seamless compatibility with SailPoint’s APIs and architecture for real-world applicability.

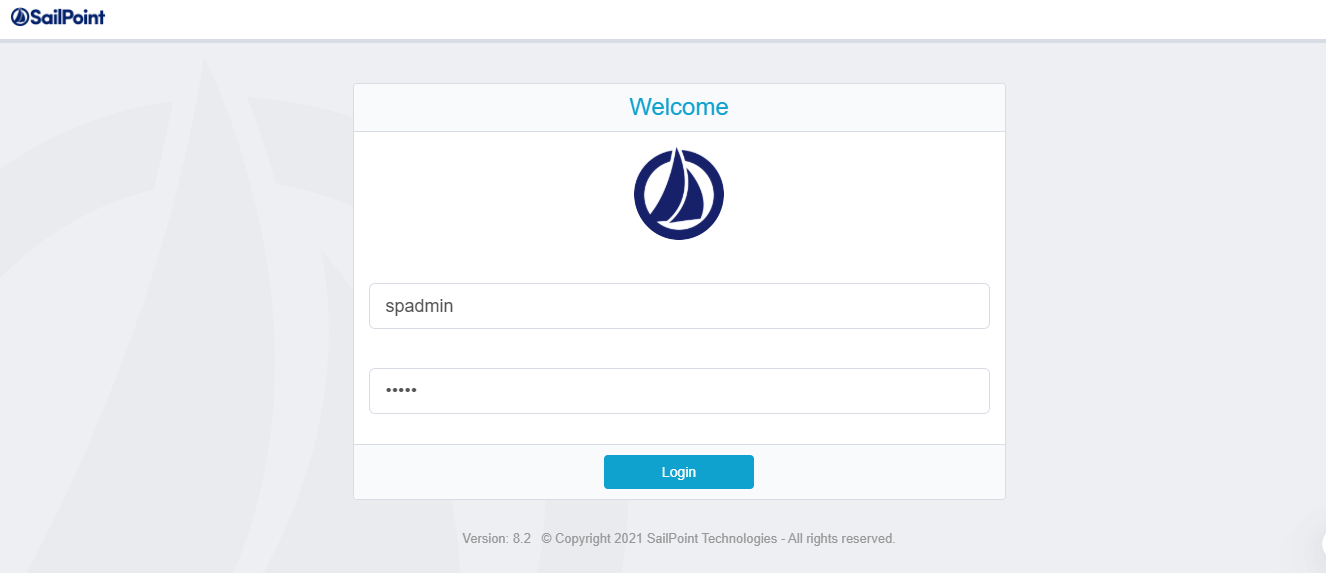
# Methodology

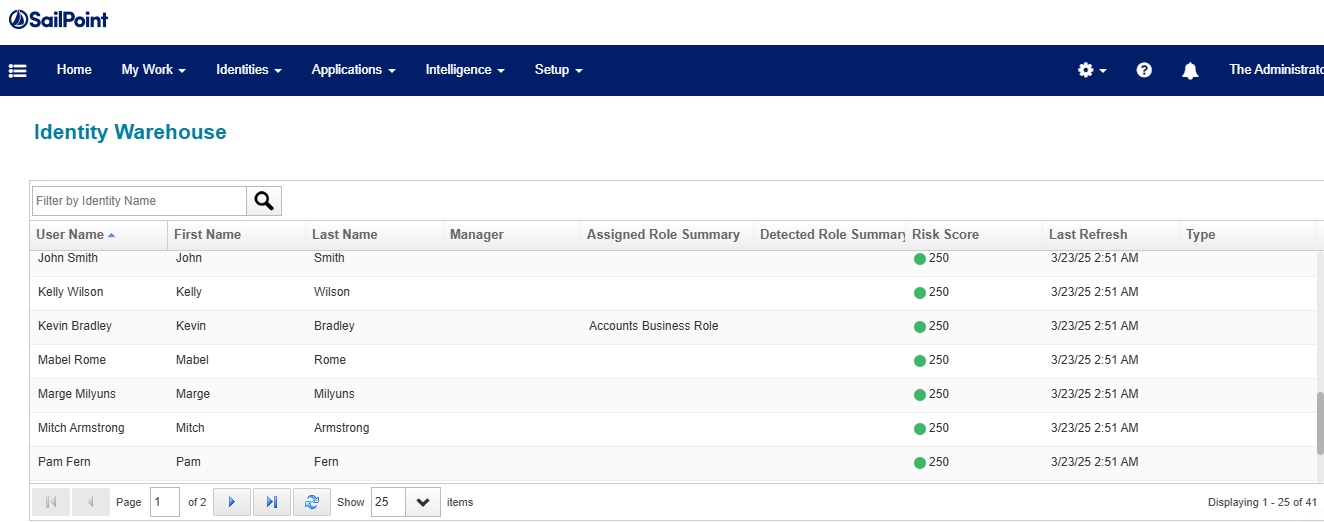
This section details the end-to-end process followed in designing and implementing a machine learning-powered solution for Intelligent Role Mining and Optimization in IAM (Identity and Access Management) within SailPoint IdentityIQ. The methodology involves data extraction, preprocessing, role mining, anomaly detection, and dashboard development.

#### 5.1 Data Extraction & Preprocessing

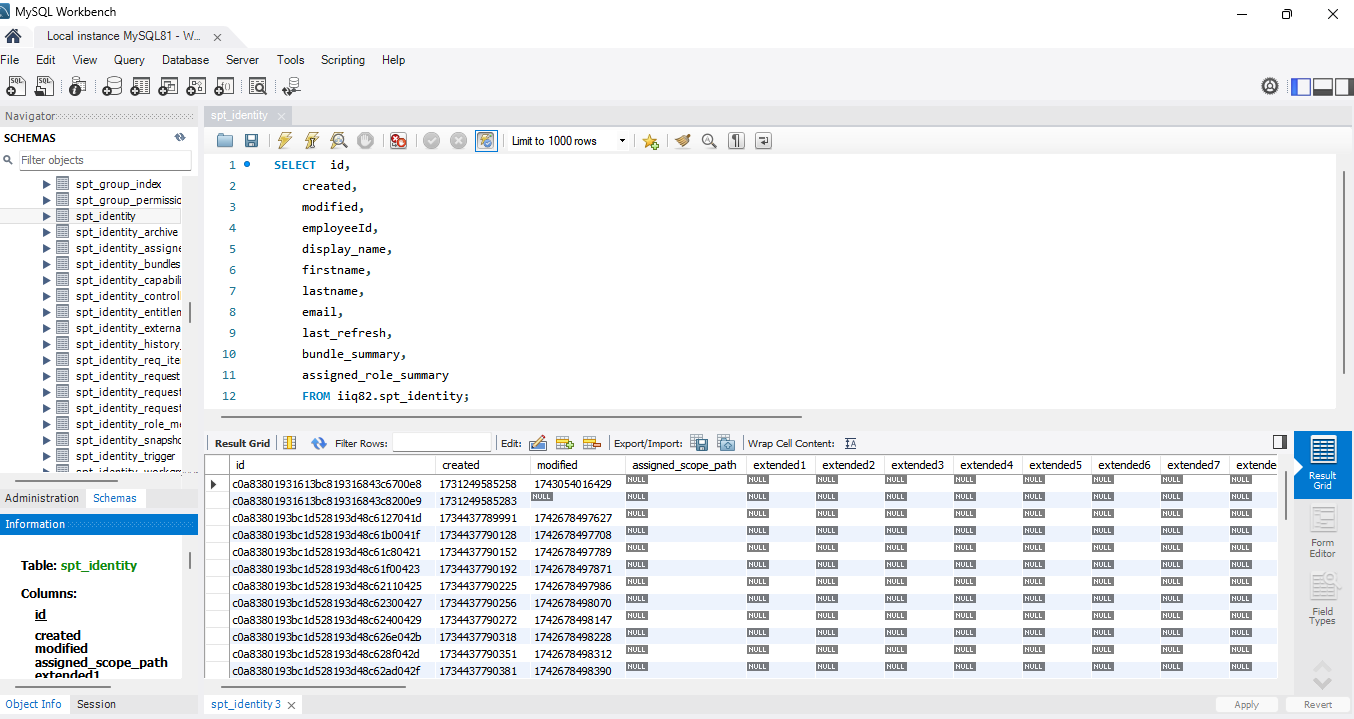
##### 5.1.1 Data Extraction from SailPoint IdentityIQ

The first step was to extract raw access data from the SailPoint IdentityIQ database. The following approaches were used:

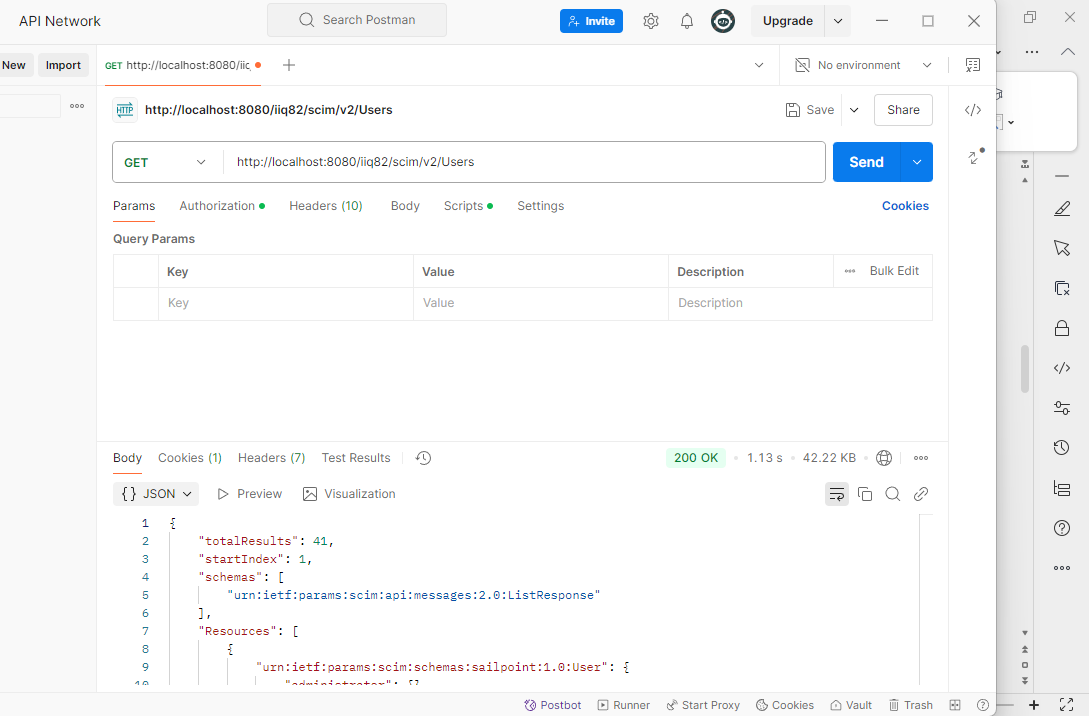






**Direct Database Queries:** SQL queries were executed on the Identity Warehouse (IW) tables to retrieve information about users, assigned roles, entitlements, and permissions.

**API-Based Extraction:** REST API endpoints provided by SailPoint IdentityIQ were utilized to fetch real-time access data, ensuring the extracted data remained up to date(Postman).



**Log Data Retrieval:** Access logs were analyzed to track historical role assignments and modifications.

**Dataset Used:** The dataset was loaded from a GitHub repository containing access control detailshttps://raw.githubusercontent.com/anilgovind994/Role-Mining-in-SP-using-Ml/refs/heads/main/Data%20-%20Copy.csv

.

##### 5.1.2 Data Preprocessing & Transformation

The extracted data was preprocessed to ensure it was clean and suitable for machine learning models. The following preprocessing steps were undertaken:

**Handling Missing & Inconsistent Data:** Missing values in user roles, entitlements, and access patterns were identified and handled appropriately.

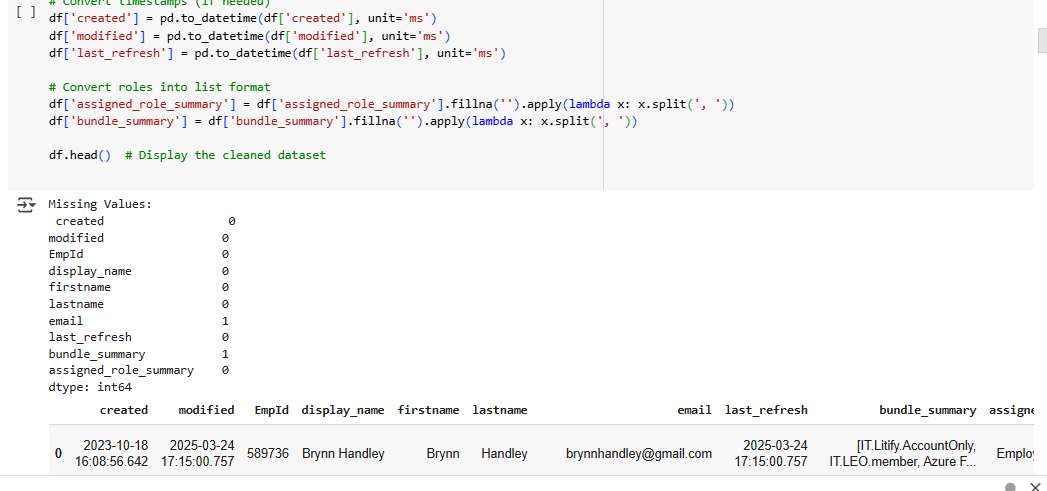
**Data Normalization:** Access attributes (e.g., role assignments, user department, last login time) were normalized to a common format to improve model accuracy.

**Duplicate Removal:** Redundant records arising from API calls or database queries were removed to avoid bias in clustering and anomaly detection models.

**Feature Engineering:**

**Role-Based Access (RBA) Vectors:** Using MultiLabelBinarizer, access permissions were converted into numerical representations to facilitate machine learning analysis.

**User Behavior Metrics:** Additional attributes such as access frequency, last access date, and department-specific access patterns were created.



#### 5.2 Role Mining & Optimization

##### 5.2.1 Clustering for Role Discovery

To optimize access control, a role-mining approach was implemented using clustering techniques:

**K-Means Clustering:** Grouped users with similar access patterns into clusters to identify redundant and efficient roles.

**Hierarchical Clustering:** Used to detect role hierarchies and merge closely related roles.

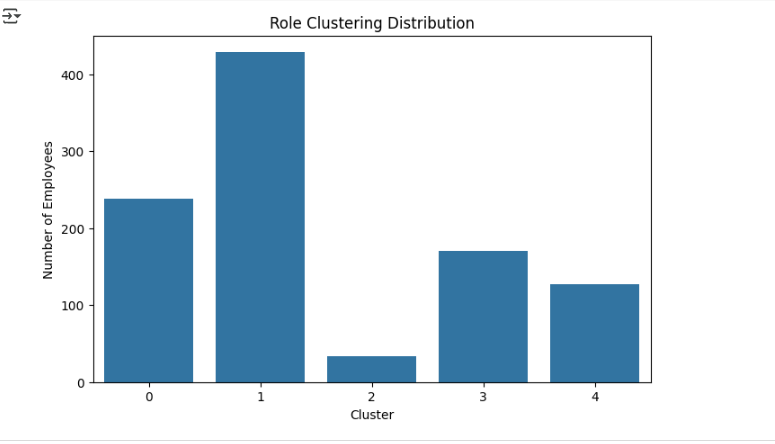
**DBSCAN (Density-Based Clustering):** Helped detect noise points, i.e., users who did not fit into existing role structures.

##### 5.2.2 Role Optimization Framework

**Merged Redundant Roles:** Closely related roles with minimal permission differences were combined to reduce unnecessary complexity.

**Suggested New Roles:** If a group of users exhibited a similar but previously undefined access pattern, a new role was suggested.

**Fine-Tuning of Role Assignments:** Analyzed and removed toxic combinations (e.g., conflicting privileges) to enhance security.



#### 5.3 Anomaly Detection in Access Patterns

##### 5.3.1 Anomaly Detection Model Selection

Anomaly detection was implemented to identify unusual access behaviors using:

**Isolation Forest:** Identified outliers based on how easily a user’s access pattern could be isolated from the majority.

**DBSCAN (Density-Based Spatial Clustering of Applications with Noise):** Detected users whose access patterns were significantly different from standard clusters.

**Autoencoders (Deep Learning-based):** Encoded normal user access behavior and flagged deviations.

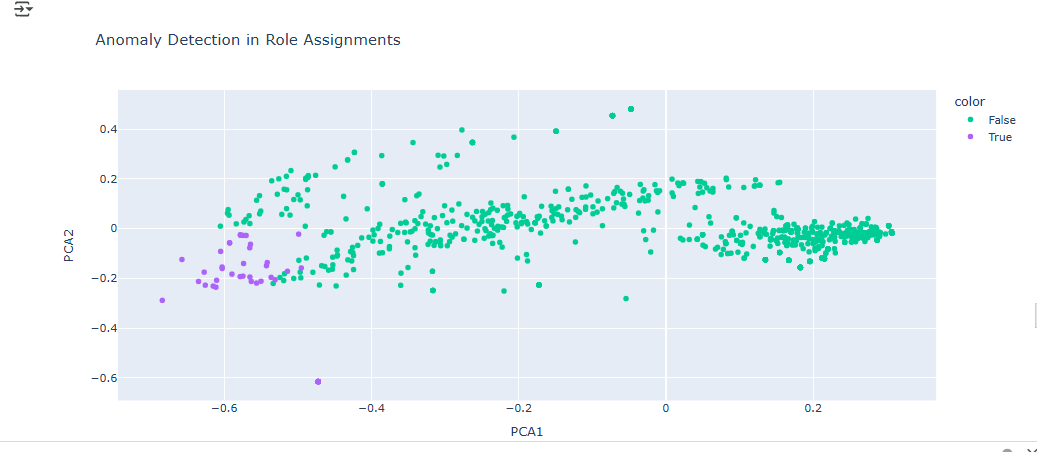
##### 5.3.2 Anomaly Classification

Once anomalies were detected, they were classified into:

**Unauthorized Access Attempts:** Users trying to access systems they were not assigned to.

**Privilege Escalation Risks:** Sudden access to high-privilege roles without proper authorization.

**Inactive or Dormant Accounts with Active Access:** Identified accounts that were no longer in use but still had permissions.





#### 5.4 Dashboard Development & Visualization

##### 5.4.1 Dashboard Overview

An interactive dashboard was developed using Dash & Plotly, providing the following visualizations used Google Colab to perform the Machine Learning on User Roles.

https://colab.research.google.com/drive/15eDwF0hIk0mMOrqeGa\_xNpaLO0nosqTH#scrollTo=RfVB1g6aC0R7

**Cluster Visualization:** A scatter plot displaying clustered roles, helping administrators understand role groupings.

**Anomaly Overview:** A table listing flagged users, their anomaly scores, and the nature of detected anomalies.

**Role Optimization Table:** A tabular representation of optimized role recommendations with original and suggested role mappings.

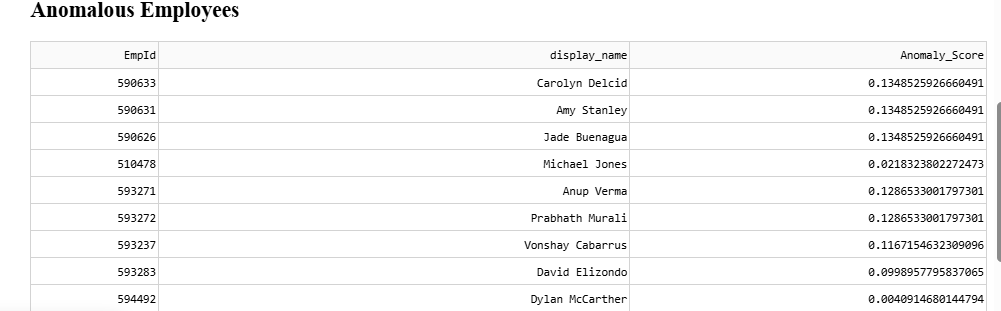
##### 5.4.2 Key Features

**Filter & Search Options:** Allowed admins to explore specific users, roles, or anomaly categories.

**Graphical Insights:** Role clusters and anomaly trends were visualized for better decision-making.

**Export Functionality:** Enabled downloading optimized role recommendations for further analysis.





#### 5.5 Validation & Testing

To ensure the accuracy and efficiency of the solution:

**Cross-Validation** was performed on clustering and anomaly detection models.

**Comparison with Manual Role Audits** to verify if the ML-based role suggestions aligned with security best practices.

**Security Testing** to prevent unauthorized modifications to access roles.

### Summary of the Methodology

This methodology ensures a structured, data-driven approach to role mining and anomaly detection in SailPoint IdentityIQ. By automating access governance through machine learning, the solution enhances security, simplifies role management, and provides actionable insights for IAM administrators.

# Plan of Work

|  |  |  |
| --- | --- | --- |
| **Phases** | **Start Date-End Date** | **Work to be done** |
| Dissertation Outline | 25 Jan 2025 – 05 Feb 2025 | Literature Review and preparation of Dissertation Outline |
| Design & Development | 06 Feb 2025 – 10 Mar 2025 | Implementation of role mining and anomaly detection models. |
| Testing | 11 Mar 2025 – 20 Mar 2025 | Testing machine learning models and integrating them with SailPoint. |
| Dissertation Review | 21 Mar 2025 – 30 Mar 2025 | Submission of the draft dissertation for supervisor feedback and revisions. |
| Submission | 01 Apr 2025 | Final Review and submission of Dissertation |

# Literature References

1. *The project builds on existing research and practices in IAM and SailPoint. Key references include:*
2. *1.SailPoint Technologies. "Advanced Role Mining and Access Governance," SailPoint Documentation, 2023.*
3. *2.K. Vaidya, A. Atluri. "Role Optimization in Identity Management Systems," IEEE Transactions, 2022.*
4. *3.S. Brown. "AI Applications in SailPoint IAM," AI Security Journal, 2021.*
5. *4.R. Chandramouli. "Role-Based Access Control for Enhanced Security," Journal of Cybersecurity Innovations, 2023.*
6. *5.L. Chen. "Visualization and Analytics in IAM Systems," Visualization in Computing, 2021.*
7. *6.M. Zhou. "Anomaly Detection in SailPoint IAM Using Machine Learning," ACM Transactions on Information and System Security, 2020.J. R. Trippas, D. Spina, L. Cavedon, H. Joho, and M. Sanderson. Informing the design of spoken conversational search: Perspective paper. In Proceedings of the 2018 Conference on Human Information Interaction & Retrieval, CHIIR ’18, pages 32–41, New York, NY, USA, 2018. ACM.*
8. *J. R. Trippas, D. Spina, L. Cavedon, and M. Sanderson. How do people interact in conversational speech-only search tasks: A preliminary analysis. In Proceedings of the 2017 Conference on Conference Human Information*Interaction and Retrieval, CHIIR ’17, pages 325–328, New York, NY, USA, 2017. ACM.

# Particulars of the Supervisor and Examiner

|  |  |  |
| --- | --- | --- |
|  | **Supervisor** | **Additional Examiner** |
| Name |  |  |
| Qualification |  |  |
| Designation |  |  |
| Employing Organization and Location |  |  |
| Phone No.(with STD Code) |  |  |
| Email Address |  |  |

# Remarks of the Supervisor

The study and project seem to very promising. The student had covered all the relevant literature review. The project is holding the current domain of research and justified the level of dissertation for master studies. The outcome covered by the project seems very interesting and include all the state of arts in today’s era of information science. The project also includes the user study on the working prototype. This activity completely validates the research project. This project also contains the prospect of publication in future. After the initial meeting with the student, I got an impression that she is hard working student and have the skills to implement the experiment in standard research paradigm. I approved the following project as the supervisor.

**Information about the Supervisor:**

**BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI**

**WORK INTEGRATED LEARNING PROGRAMMES (WILP) DIVISION**

**SECOND SEMESTER OF ACADEMIC YEAR 2021-2022**

**(INSERT COURSE NUMBER HERE) : (INSERT COURSE TITLE HERE) OUTLINE**

|  |  |
| --- | --- |
| **STUDENT ID No.** |  |
| **NAME OF THE STUDENT** |  |
| **STUDENT'S EMAIL ADDRESS** |  |
| **STUDENT’S EMPLOYING ORGANIZATION & LOCATION** |  |
| **SUPERVISOR’S NAME** |  |
| **SUPERVISOR’S EMPLOYING ORGANIZATION & LOCATION** |  |
| **SUPERVISOR’S EMAIL ADDRESS** |  |
| **ADDITIONAL EXAMINAER’S NAME** |  |
| **ADDITIONAL EXAMINER’S EMPLOYING ORGANIZATION & LOCATION** |  |
| **ADDITIONAL EXAMINER’S EMAIL ADDRESS** |  |
| **DISSERTATION / PROJECT / PROJECT WORK TITLE** |  |

Please prepare the outline as a separate document with the following sections along with the above identification information.

1. Cover Page with Student ID No., Name, Course Number, Course Title and Dissertation / Project / Project Work Title, Broad Academic Area of Work.

2. Background (Relevance of the Project to the current work environment in the employing organization)

3. Objectives

4. Scope of Work (To be done by the student independently)

5. Plan of Work (Work to be done during the semester)

6. Literature References

7. Particulars of the Supervisor and Additional Examiner

8. Remarks of the Supervisor

|  |  |  |
| --- | --- | --- |
|  |  |  |
| **Signature of Student** | **Signature of Supervisor** | **Signature of Additional Examiner** |
| **Name:** | **Name:** | **Name:** |