A REPORT
ON

DATA DRIVEN SOLUTION FOR CONNECTING EDUCATION AND EMPLOYMENT

Submitted by,

Mr. DHANUSH C R - 20211CSE0671

Under the guidance of,

Mr. SYED MOHSIN ABBASI

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

At

GAIN MORE KNOWLEDGE
REACH GREATER HEIGHTS

PRESIDENCY UNIVERSITY

BENGALURU

MAY 2025

PRESIDENCY UNIVERSITY

PRESIDENCY SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the Internship/Project report **"DATA DRIVEN SOLUTION FOR CONNECTING EDUCATION AND EMPLOYMENT"** being submitted by "DHANUSH CR" bearing roll number "20211CSE0671" in partial fulfillment of the requirement for the award of the degree of Bachelor of Technology in Computer Science and Engineering is a bonafide work carried out under my supervision.

Mr. SYED MOHISN ABBASI
ASSISTANT PROFESSOR
PSCS / PSIS
Presidency University

Dr. ASIF MOHAMED H B
ASSOCIATE PROFESSOR & HoD
PSCS
Presidency University

Dr. MYDHILI NAIR
Associate Dean
PSCS
Presidency University

Dr. SAMEERUDDIN KHAN
Pro-Vice Chancellor - Engineering
Dean –PSCS / PSIS
Presidency University

PRESIDENCY UNIVERSITY

PRESIDENCY SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

DECLARATION

I hereby declare that the work, which is being presented in the report entitled **"DATA DRIVEN SOLUTION FOR CONNECTING EDUCATION AND EMPLOYMENT"** in partial fulfillment for the award of Degree of **Bachelor of Technology** in **Computer Science and Engineering**, is a record of my own investigations carried under the guidance of **Mr. SYED MOHSIN ABBASI, ASSISTANT PROFESSOR, Presidency School of Computer Science and Engineering, Presidency University, Bengaluru.**

I have not submitted the matter presented in this report anywhere for the award of any other Degree.

DHANUSH CR – 20211CSE0671

INTERNSHIP COMPLETION CERTIFICATE

UPTOSKILLS
Skills for lifetime!

CERTIFICATE
OF APPRECIATION

Proudly presented to :

DHANUSH C R

In recognition of his hard work and dedication in completing the internship as a **Data analytics intern** from 22/01/2025 to 22/04/2025 at UptoSkills Company.

UPTOSKILLS Auth Sign Delhi

Shivam Agrawal
Mentor

This certificate can be verified at hr@uptoskills.com

iv

ABSTRACT

In today's dynamic job market, a significant disconnect exists between the skills taught in academic institutions and those required by industries. This project addresses the growing concern of graduate unemployability by leveraging data analytics to bridge the gap between education and employment. Conducted as a Data Analyst internship project, it utilizes real-time data mining and analysis to identify the roles companies are hiring for and map those needs against college-level training opportunities.

The project was divided into two key phases:

1. Data Mining: Collected and categorized information from multiple online sources such as LinkedIn, Internshala, Naukri, and others to gather job and internship postings across Indian cities. Simultaneously, Training and Placement Officer (TPO) details were collected from engineering colleges in the same regions to build a comprehensive dataset.
2. Data Analysis: Cleaned and structured the company data using tools like Excel, Power BI, and Microsoft Azure (ADLS, ADB). The analysis involved creating dashboards, removing duplicates, transforming columns, and visualizing hiring patterns across job roles, locations, and qualification criteria.

The core insights enabled the identification of high-demand roles and skills region-wise. These insights can be used by the company to:

- Provide targeted training programs to colleges based on in-demand roles
- Connect with companies to offer skilled candidates, acting as a recruitment partner
- Facilitate corporate training based on gaps in skill readiness
- Leverage TPO contact data to expand outreach to educational institutions

By acting as a data-driven bridge between colleges and companies, the organization not only enhances placement opportunities for students but also generates dual revenue streams—through training services for colleges and recruitment services for companies.

v

ACKNOWLEDGEMENTS

First of all, we indebted to the **GOD ALMIGHTY** for giving me an opportunity to excel in our efforts to complete this project on time.

We express our sincere thanks to our respected dean **Dr. Md. Sameeruddin Khan**, Pro-VC - Engineering and Dean, Presidency School of Computer Science and Engineering & Presidency School of Information Science, Presidency University for getting us permission to undergo the project.

We express our heartfelt gratitude to our beloved Associate Dean **Dr. Mydhili Nair,** Presidency School of Computer Science and Engineering, Presidency University, and **Dr. Asif Mohamed H B**, Head of the Department, Presidency School of Computer Science and Engineering, Presidency University, for rendering timely help in completing this project successfully.

We are greatly indebted to our guide **Mr. Syed Mohsin Abbasi, Assistant Professor** and Reviewer **Dr. Jayavadival Ravi, Associate Professor**, Presidency School of Computer Science and Engineering, Presidency University for his inspirational guidance, and valuable suggestions and for providing us a chance to express our technical capabilities in every respect for the completion of the internship work.

We would like to convey our gratitude and heartfelt thanks to the CSE7301 Internship/University Project Coordinator **Mr. Md Ziaur Rahman and Dr. Sampath A K,** department Project Coordinators **Dr. Jayanthi K** and Git hub coordinator **Mr. Muthuraj.**

We thank our family and friends for the strong support and inspiration they have provided us in bringing out this project.

DHANUSH CR

LIST OF TABLES

LIST OF FIGURES

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

1.1 Understanding the Education-to-Employment Ecosystem

In the evolving landscape of global job markets, the alignment between educational institutions and industry requirements is increasingly under scrutiny. Educational systems are designed to impart knowledge, but the dynamic nature of industry demands often leaves a noticeable gap between what students learn and what employers expect. This gap can lead to low employability rates despite high educational qualifications, ultimately affecting both students' career prospects and company productivity.

1.1.1 The Critical Role of Campus-Industry Linkages

Campus-Industry linkages play a vital role in ensuring that students are equipped with job-relevant skills. Training and Placement Officers (TPOs) act as the communication bridge between academic institutions and companies, facilitating internships, placements, and skill development programs. A well-connected ecosystem, where real-time hiring data informs training initiatives, can drastically enhance student employability and institutional placement records.

1.1.2 Growing Demand for Data-Driven Interventions

With access to massive volumes of employment and academic data, there is a growing opportunity to use data analytics to inform and guide skill-building strategies. By analyzing hiring patterns, in-demand job roles, and regional employment trends, institutions can make informed decisions about what skills to teach and which companies to approach for recruitment. Similarly, companies can benefit from a pre-vetted talent pipeline trained in their specific requirements.

1.2 Challenges in the Current Campus-to-Corporate Model

While colleges and companies independently pursue growth and development, their collaboration remains inconsistent and inefficient due to various systemic challenges.

1.2.1 Mismatch Between Curriculum and Industry Needs

One of the primary issues is the lack of industry-aligned curricula in colleges. Syllabi are often outdated or theoretical, with little focus on emerging technologies or soft skills essential in a corporate environment. This mismatch delays job readiness and requires additional training from the employers' side.

1.2.2 Unstructured Placement Ecosystem

Most placement ecosystems are run based on personal contacts and traditional processes. There is a lack of standardized systems for identifying in-demand roles, mapping them to student competencies, or tracking regional hiring trends. This leads to inefficiencies and missed opportunities for both students and companies.

1.2.3 Inconsistent Communication Channels

TPOs often struggle to maintain consistent communication with potential recruiters due to the absence of a centralized database or structured outreach strategy. Similarly, companies find it hard to identify which institutions are actively training students in specific skillsets, causing friction in the recruitment process.

1.3 Role of Data Analytics in Solving the Placement Gap

The application of data analytics can significantly streamline the education-to-employment pipeline. By using collected data effectively, decision-makers can identify key patterns, opportunities, and mismatches in the system.

1.3.1 Regional Hiring Insights

A city-level or regional analysis of hiring trends allows institutions to understand the most in-demand skills and roles in the job market. Although students may move across cities for jobs, understanding regional trends helps tailor localized training programs that are scalable and flexible.

1.3.2 Profiling TPOs and Academic Institutions

Compiling and analysing TPO data across engineering colleges in India creates a strategic outreach opportunity. This enables centralized communication and collaboration with institutions based on their readiness, student strength, and willingness to adopt industry-relevant training programs.

1.3.3 Company-Wise Job Role Forecasting

By mining job postings, internships, and hiring campaigns from a wide range of companies, the data can reveal high-demand job roles, required skills, and seasonal hiring patterns. These insights help bridge the timing and relevance gap between training delivery and actual hiring cycles.

1.4 Strategic Opportunities for Industry and Academia

This project not only solves a problem—it opens new revenue and collaboration models for all stakeholders involved.

1.4.1 For the Training Organization (Internship Company)

The organization where the data analyst internship is conducted can position itself as the central bridge between academia and industry. With access to TPO and hiring data, the company can:

- Offer skill-based training programs to colleges.
- Help companies hire pre-trained, job-ready candidates.
- Receive commissions from both colleges and companies for successful placements and training services.

1.4.2 For Colleges and Universities

Colleges benefit from increased placement rates, better brand reputation, and stronger industry partnerships. With guidance on trending skills and targeted training modules, students are better prepared for job interviews and industry expectations.

1.4.3 For Companies and Recruiters

Companies reduce their hiring costs and time by accessing a pipeline of candidates trained specifically in their requirements. Corporate training programs can also be offered to upskill freshers and existing employees, making onboarding and role transitions smoother.

1.5 Scope of the Project Work

The project's scope is comprehensive, touching all major stakeholders in the education and employment ecosystem.

1.5.1 Data Mining and Analysis

During the internship, the project involved collecting structured data from engineering institutions and corporate hiring portals. The first phase focused on mining TPO information and company job listings. The second phase involved cleaning, visualizing, and analysing this data using tools like Excel, Power BI, PySpark, and Azure-based cloud platforms.

1.5.2 Insight Generation and Mapping

The insights derived from the data helped in identifying:

- Gaps between academic training and company requirements.
- High-demand job roles across various cities.
- Regional skill trends.

- Opportunities for new training programs and partnerships.

1.5.3 Business Strategy Recommendation

Finally, a go-to-market model was conceptualized for the company to scale this solution and turn it into a sustainable, revenue-generating offering. The strategy includes monetizing both college-side and company-side services using the data-driven insights.

1.6 Significance of the Project

This project stands as a proof-of-concept for how data analytics can revolutionize the education-employment pipeline.

1.6.1 Driving Employability and Opportunity

By connecting students with relevant, high-demand job roles and ensuring they are trained accordingly, the project boosts employability and reduces the job-search burden.

1.6.2 Empowering Decision-Makers

TPOs, recruiters, and company executives can make smarter, faster, and more informed decisions about collaborations, training investments, and hiring strategies.

1.6.3 Sustainable Value Creation

By aligning all stakeholders through data, the project facilitates continuous learning, adaptability, and long-term impact in the skill-development ecosystem.

CONCLUSION

The disconnect between educational outcomes and industry requirements has long been a bottleneck in the smooth transition of students from college to career. This project takes a critical step in addressing this challenge by leveraging data analytics to uncover actionable insights about institutional readiness, company hiring trends, and regional skill demand. By systematically collecting and analysing data on TPOs, engineering colleges, and job listings across India, the project aims to build a scalable bridge that brings companies and colleges into direct, mutually beneficial collaboration.

The introduction chapter has explored the current gaps in the campus-to-corporate pipeline, emphasized the value of data-driven strategies in overcoming these gaps, and outlined the project’s vision for sustainable impact. It has also highlighted how each stakeholder—students, colleges, companies, and training organizations—stands to benefit from this integrated approach.

Chapter 2

LITERATURE SURVEY

The primary aim of this project is to leverage data-driven technologies to bridge the gap between education and employment by aligning academic training with evolving industry demands. To lay the groundwork for this initiative, it is essential to examine existing literature on the integration of data analytics in educational planning and workforce development. This chapter provides a structured review of research studies, tools, and methodologies that explore how big data, predictive analytics, cloud platforms, and visualization technologies have been applied to identify skill gaps, inform training programs, and improve employment outcomes. The insights from this literature form the foundation upon which the proposed system is conceptualized and developed.

2.1 Overview of Data Analytics in Education and Employment

The dynamic shifts in the global economy, accelerated by digitization and automation, have necessitated a stronger alignment between the skills taught in educational institutions and the evolving needs of the job market. Literature in this domain emphasizes the growing significance of data-driven strategies to bridge this educational-employment gap. Traditional educational frameworks often fail to keep pace with rapid industry innovations, resulting in graduates who are theoretically equipped but practically unprepared for employment.
Modern data analytics allows stakeholders—including policymakers, educators, and recruiters—to analyse patterns in educational outcomes, job requirements, student interests, and market trends. By extracting insights from large datasets such as institutional records, alumni data, and job listings, a more holistic understanding of talent readiness can be developed.

2.1.1 Evolution of Data-Driven Decision Making

• Early efforts focused primarily on academic metrics like grades, dropout rates, and enrolment statistics to evaluate student success.
• Employment trend analysis traditionally relied on static survey data and government labour reports, often outdated by the time of publication.
• Contemporary approaches use real-time analytics sourced from online platforms, corporate job boards, and university records to provide dynamic insights.

• Big data and AI tools enable institutions to shift from reactive to predictive models of decision-making.

2.1.2 The Urgent Need for Alignment

• Several studies, including those by McKinsey and NASSCOM, indicate a stark mismatch between what students learn and what companies require.

• Data-driven systems allow curriculum development based on the actual demand for skills like cloud computing, data science, and soft skills.

• Institutions are now exploring frameworks where course offerings are adjusted periodically based on insights from job market analytics.

2.2 Data Mining Techniques for Talent Mapping

Effective talent mapping begins with accurate and exhaustive data collection. Literature points to web scraping and automated data mining as key techniques in gathering institutional and job market information at scale. The ability to extract contact data, course offerings, recruitment details, and company profiles forms the foundation for deeper analysis.

2.2.1 Techniques for Web Data Extraction

• Web scraping tools such as Beautiful Soup, Selenium, and Scrapy are frequently cited for gathering real-time data from educational and corporate websites.

• APIs from platforms like LinkedIn, Naukri.com, and AICTE offer structured data access but may come with limitations in scope and cost.

• Unstructured formats—like PDF brochures and non-standard HTML layouts—pose significant challenges to consistent extraction.

2.2.2 Normalization and Preprocessing

• Data cleaning and transformation processes are required to bring heterogenous datasets into a unified schema.

• NLP methods such as Named Entity Recognition (NER) are used to identify colleges, companies, skills, locations, and job roles from raw text.

• Standardization of naming conventions, deduplication, and removal of outliers are crucial before meaningful insights can be derived.

2.2.3 Real-World Applications

• Several job matching portals use backend scraping and parsing tools to compile job descriptions, extract keywords, and recommend candidate profiles.

• Universities such as IITs and IIMs are increasingly collaborating with EdTech startups to leverage data for placement preparation and recruiter engagement strategies.

2.3 Cloud-Based Data Handling and Analysis

Cloud computing has revolutionized the way educational and employment data is stored, processed, and shared. Among several platforms, Microsoft Azure stands out in academic literature due to its comprehensive suite of tools supporting large-scale data pipelines.

2.3.1 Capabilities of Microsoft Azure for Educational Analytics

• Azure Data Lake Storage (ADLS) allows hierarchical data storage that scales seamlessly for varied formats—CSV, JSON, Parquet, and more.

• Azure Databricks (ADB), integrated with Apache Spark, enables parallel processing of large datasets, especially for real-time analytics.

• PySpark provides an accessible interface for writing scalable data transformation and machine learning tasks.

2.3.2 Benefits of Cloud Platforms in Research

• Centralized access to datasets improves collaboration among educators, researchers, and placement cells.

• Literature confirms that cloud platforms reduce dependency on local infrastructure, making them ideal for large-scale educational research.

• Security and role-based access controls ensure that sensitive data (e.g., student records or company partnerships) is protected.

2.3.3 Challenges in Cloud Adoption

• Smaller colleges and Tier-3 institutions often lack the digital infrastructure and trained manpower to implement such solutions.

• Concerns around privacy, especially in compliance with laws like GDPR and India’s PDP Bill, remain major barriers to full-scale implementation.

• Misconfiguration in access policies can lead to data breaches, which literature strongly cautions against.

2.4 Predictive Analytics in Skill Gap Analysis

Skill gap analysis refers to identifying the disparity between the capabilities of job seekers and the requirements of employers. Predictive analytics, powered by machine learning, is increasingly being used to recommend career paths, required upskilling, and job matches.

2.4.1 Algorithms for Career Forecasting

• Classification algorithms can segment students based on career readiness.

• Regression models predict employability scores based on past academic and non-academic metrics.

• Recommendation systems match students with internships, online courses, or job roles based on personal and market data.

2.4.2 Key Findings from Recent Studies

• According to a study by Infosys, data-driven upskilling suggestions result in a 22% higher job placement rate among engineering graduates.

• Coursera’s 2023 report notes that learners who follow AI-generated career paths complete certifications 1.7x faster than others.

• AICTE’s AI-based model for curriculum modernization uses real-time job data to guide updates in technical education programs.

2.4.3 Ethical and Technical Challenges

• Predictive models can inherit bias from historical data, disadvantaging certain groups.

• Model interpretability is a concern in high-stakes environments like college placements.

• Frequent retraining of models is required to keep up with dynamic job market trends.

2.5 Visualization and Stakeholder Communication

Data visualization plays a crucial role in translating complex analytical insights into actionable knowledge for stakeholders such as students, placement officers, academic councils, and industry partners.

2.5.1 Importance of Visual Communication

• Power BI, Tableau, and Google Data Studio are widely used for creating dashboards that display metrics like placement ratios, sectoral trends, and recruiter activity.

• Custom dashboards allow TPOs to monitor training effectiveness and identify student groups that require additional support.

• Policy makers use visualizations to identify underperforming departments or regions.

2.5.2 Case Studies

• Institutions like VIT, SRM, and BITS Pilani have adopted dashboard systems to guide campus recruitment strategies.

• EdTech platforms often include dashboard features for students to track their own progress, certifications, and job readiness.

2.5.3 Common Pitfalls

• Misleading visuals due to poor chart selection or scaling can distort interpretation.

• Lack of training in data storytelling may limit the impact of even the most accurate insights.

• Literature suggests including interactivity (e.g., filters, drill-downs) to enhance decision-making.

CONCLUSION

The literature clearly outlines a growing convergence of education and employment ecosystems through the application of advanced data analytics, machine learning, and cloud-based platforms. From web scraping for data mining to using predictive analytics for skill mapping, each technological component contributes to solving the long-standing disconnect between what is taught in colleges and what is demanded in the job market. Microsoft Azure’s ecosystem, particularly ADLS and Databricks, enables large-scale data processing, while visualization tools like Power BI empower non-technical stakeholders to participate in data-driven decision-making. However, challenges such as data inconsistency, infrastructural limitations, and privacy concerns must be systematically addressed. As the reviewed literature suggests, the future lies in integrated platforms that facilitate dynamic, responsive, and inclusive education-to-employment pathways.

Chapter 3

RESEARCH GAPS OF EXISTING METHODS

The current landscape of education-to-employment systems has seen rapid integration of technology and data analytics. Despite several initiatives to connect academic training with industry demands, a disconnect still persists between educational outputs and employment requirements. This chapter outlines the key research gaps identified through a critical review of existing systems, models, and frameworks, focusing on the areas of data collection, curriculum alignment, employment forecasting, and technology adoption.

3.1 Research Gaps in Data Collection and Integration

Effective decision-making in education-employment alignment depends heavily on robust and reliable data. However, several limitations hinder comprehensive and actionable insights.

3.1.1 Fragmented Data Sources

- Most data related to student performance, course offerings, placement outcomes, and job openings are stored across disparate platforms with little to no integration.
- There is no centralized system that allows seamless access to stakeholders from academia and industry.
- Gap: A unified data infrastructure that consolidates educational and employment data is largely missing in current approaches.

3.1.2 Lack of Real-Time Data Analytics

- Current systems rely on periodic data updates, often quarterly or annually, leading to lagged insights.
- The delay impacts curriculum revisions, recruitment planning, and student preparation.
- Gap: Integration of real-time data streams and dashboards to allow timely interventions remains an under-researched area.

3.1.3 Limited Use of Predictive Data Models

- Traditional reporting tools are descriptive in nature and do not support forecasting or pattern recognition for future job trends.
- Gap: The development of AI-powered predictive models that correlate student skills and academic outcomes with future employment scenarios needs further research.

3.2 Research Gaps in Curriculum and Skill Alignment

Despite repeated calls for industry-academic collaboration, existing educational structures often fail to keep pace with evolving industry demands.

3.2.1 Static Curriculum Structures

- Many institutions follow rigid academic curriculums with limited scope for dynamic updates based on job market trends.
- Gap: There is a lack of frameworks to ensure agile, data-driven curriculum evolution that responds to emerging technologies and employer expectations.

3.2.2 Skill Mismatch Between Graduates and Employers

- Research indicates that graduates often lack the practical skills and soft competencies expected in real-world job roles.
- Gap: Existing models do not adequately address the granular mapping of job descriptions to course content and learning outcomes.

3.2.3 Inadequate Feedback Loops from Employers to Institutions

- Communication between recruiters and academic institutions is often limited to placement seasons.
- Gap: Continuous, structured feedback systems that help institutions refine their programs are either informal or non-existent in current setups.

3.3 Research Gaps in Employment Forecasting

Accurate prediction of future employment needs is crucial for preparing students effectively, yet most existing models lack depth and foresight.

3.3.1 Insufficient Granularity in Labor Market Forecasts

- National or regional employment data lacks specificity, making it difficult to apply insights to local institutions or niche sectors.
- Gap: Research into micro-level employment forecasting models that cater to specific industries, geographies, or educational streams is limited.

3.3.2 Weak Integration of Industry Trends in Educational Planning

- Sectors like AI, data science, green tech, and digital finance are evolving rapidly, but academic planning often fails to anticipate and reflect such growth.
- Gap: Mechanisms to translate emerging industry trends into academic decisions are either manual or reactive rather than proactive and automated.

3.3.3 Limited Scenario-Based Simulations

- Institutions rarely use scenario modelling to test how different skill training combinations affect employability outcomes.
- Gap: The application of simulation tools for curriculum planning and placement forecasting is still in a nascent stage.

3.4 Research Gaps in Technology Implementation

While digital platforms and cloud-based systems are being increasingly adopted, several barriers limit their effectiveness in bridging education-employment gaps.

3.4.1 Lack of End-to-End Automation in the Hiring Pipeline

- Many institutions use technology for placement registration or resume screening, but the full process from student training to job matching is often fragmented.
- Gap: Research into holistic platforms that automate the full lifecycle of employability—from skill development to recruitment—is insufficient.

3.4.2 Inaccessible Interfaces for Rural or Underserved Students

- Students from Tier-2 and Tier-3 cities often lack access to personalized platforms or analytics tools that can guide career development.
- Gap: Inclusion-focused technology design that accounts for diverse student needs and resource availability remains a challenge.

3.4.3 Data Privacy and Ethical Concerns in Analytics

- With growing reliance on AI and analytics, concerns regarding student data privacy and fairness in job recommendations are emerging.
- Gap: There is limited research on ethical frameworks and privacy-preserving data mining practices within the educational domain.

CONCLUSION

This chapter sheds light on the significant research gaps that exist in current methods aimed at bridging education and employment. From fragmented data systems and static curricula to underutilized forecasting models and incomplete tech adoption, there are multiple areas that need further exploration and innovation. A truly integrated, data-driven ecosystem requires not only technological upgrades but also a fundamental shift in how academia and industry collaborate, forecast, and adapt. Addressing these gaps is critical for creating a scalable, equitable, and future-ready framework that connects learning with earning.

Chapter 4

PROPOSED METHODOLOGY

The proposed methodology harnesses data-driven techniques, including Machine Learning (ML), Artificial Intelligence (AI), and cloud computing, to provide a seamless, actionable solution that bridges the gap between education and employment. The goal is to enhance employability by analysing educational programs and job market data, identifying skill gaps, and offering personalized recommendations to students, institutions, and employers. The methodology includes the following key components:

4.1 Data Collection and Integration

Data plays a central role in this solution. The system will collect and integrate data from various sources to create a comprehensive dataset for analysis.

4.1.1 Educational Data

• Information from educational institutions will be gathered, including academic curricula, skill sets, certifications offered, graduation outcomes, and placement data. • This data will be sourced from university databases, course catalogues, and other educational platforms.

4.1.2 Employment Data

• Data from job portals, employer websites, and recruitment agencies will be collected. This includes job listings, required skills, industry trends, salary information, and job demand projections. • Employment data will also be sourced from social media platforms like LinkedIn to analyse trends in job preferences and candidate qualifications.

4.1.3 Data Integration

• The collected data will be pre-processed and integrated into a central database using cloud platforms like AWS or Microsoft Azure to ensure scalability and real-time updates. • Both structured and unstructured data will be handled, enabling robust data analysis and visualization.

4.2 Skill Gap Analysis

One of the primary objectives of this methodology is to perform a detailed skill gap analysis between the skills taught by educational institutions and the skills demanded by employers.

4.2.1 Identifying Key Skills

• Using machine learning models, the system will extract key skills from both educational curricula and job descriptions. The system will map educational programs to industry requirements, identifying areas where the skills taught do not align with what is needed in the job market. • Natural Language Processing (NLP) techniques will be employed to analyse job postings and course content to detect specific skills and competencies.

4.2.2 Analysing Skill Gaps

• The methodology will analyse the difference in skills by comparing the competencies taught in educational programs with those required by employers. • Machine learning models will detect patterns and identify skills that are either underrepresented or missing in academic programs.

4.3 Predictive Analytics for Job Market Trends

To provide actionable insights, the system will utilize predictive analytics to forecast job market trends and future skill needs.

4.3.1 Job Demand Forecasting

• Time series analysis and regression models will be used to predict the demand for various job roles in the future. These predictions will be based on historical data from job markets, industry growth, and emerging technologies. • The predictive model will help students and institutions understand which fields will experience growth, allowing them to focus on the most relevant skill sets.

4.3.2 Emerging Skill Identification

• The system will also utilize AI to identify emerging skills and technologies that may become critical in the future job market, such as skills related to Artificial Intelligence, Machine Learning, Data Science, and blockchain. • Students will be provided with insights into which skills they should focus on to stay ahead of industry changes.

4.4 Personalized Recommendations and Career Pathways

The system will generate personalized recommendations for students, educational institutions, and employers based on data analysis.

4.4.1 Personalized Recommendations for Students

• Based on their academic background, interests, and career aspirations, students will receive suggestions for courses, certifications, and skill development programs. • The system will recommend job roles that match the student's profile and predict future career growth.

4.4.2 Personalized Recommendations for Educational Institutions

• The system will provide insights for institutions to align their curricula with industry needs, helping them update programs and introduce new courses based on the skills in demand. • Institutions can receive data-driven feedback on the effectiveness of their programs in preparing students for the job market.

4.4.3 Personalized Recommendations for Employers

• Employers will be given data-driven insights on the availability of skilled candidates, including potential sources for recruitment and training programs that align with their hiring needs. • The system will also provide a forecast of the future workforce trends to help employers plan their talent acquisition strategy.

4.5 Data Visualization and Reporting

The system will feature advanced data visualization tools to make complex datasets easy to understand and accessible to all stakeholders.

4.5.1 Dashboards

• Interactive dashboards will provide real-time data on skill gaps, job market trends, and personalized recommendations for students, institutions, and employers. • The dashboards will display key insights such as the demand for specific skills, job openings by region, and forecasts of emerging industries.

4.5.2 Reports

• Detailed reports will be generated for educational institutions to evaluate the alignment of their programs with the job market. • Students and employers will receive tailored reports showing potential career paths, relevant courses, and hiring trends.

4.6 Cloud-Based Architecture

To handle the large volumes of data involved and ensure scalability, the solution will be based on cloud computing platforms.

4.6.1 Cloud Storage and Processing

• Cloud platforms like AWS, Google Cloud, or Microsoft Azure will be used to store and process the large datasets securely. This will enable the system to handle the growth in data volume and complexity. • Data will be updated in real-time, ensuring that students and institutions have access to the most up-to-date information.

4.6.2 Scalability and Flexibility

• The cloud-based architecture will allow the system to scale as the volume of data increases. The infrastructure will be flexible enough to adapt to new data sources, models, and features as the system evolves.

4.7 Machine Learning and AI Models

The backbone of the system will be its machine learning and AI models, which will analyse and predict trends in both education and employment data.

4.7.1 Supervised and Unsupervised Learning

• The system will employ both supervised learning (for predictive models) and unsupervised learning (for clustering and identifying hidden patterns) to analyse the data. • Key algorithms include decision trees, random forests, and support vector machines for classification tasks, and k-means clustering for identifying patterns in large datasets.

4.7.2 Natural Language Processing (NLP)

• NLP will be used to analyse text data from job postings, educational curricula, and student profiles, extracting valuable insights and skill requirements. • Named entity recognition (NER) and topic modelling will help identify key skills, job roles, and areas of growth in both education and employment sectors.

4.8 Handling Data Privacy and Security

Given the sensitive nature of both educational and employment data, ensuring the privacy and security of user information is critical. The following strategies will be implemented:

4.8.1 Data Encryption

• All data stored in the system will be encrypted using industry-standard encryption algorithms, both at rest and in transit. This will prevent unauthorized access to sensitive data, ensuring that personal and educational information is kept secure. • Encryption protocols like AES-256 will be used for data at rest, and SSL/TLS encryption will secure data during transmission.

4.8.2 User Authentication and Authorization

• The system will include robust user authentication mechanisms to ensure that only authorized users have access to certain data. This will involve multi-factor authentication (MFA) and OAuth 2.0 protocols for secure login. • Users, such as students, educational institutions, and employers, will have role-based access, ensuring that each user can only access data relevant to their role.

4.8.3 Compliance with Data Protection Regulations

• The system will comply with global data protection regulations such as GDPR (General Data Protection Regulation) and CCPA (California Consumer Privacy Act), ensuring that users' personal data is handled in compliance with legal standards. • Regular audits will be conducted to ensure compliance with these regulations, and users will be given the option to delete or anonymize their data at any time.

4.9 System Evaluation and Feedback Loop

To ensure continuous improvement, the system will include a feedback loop where user interactions and feedback are incorporated into model retraining and system updates.

4.9.1 User Feedback Collection

• Users (students, institutions, and employers) will be able to provide feedback on the recommendations and the overall user experience.

• Feedback will be collected through surveys, direct input via the platform, and user behaviour tracking (with consent).

4.9.2 Model Retraining and Updates

• The system will include periodic updates to retrain the machine learning models using the latest data. This will ensure that the platform remains aligned with the current state of the job market and educational trends.

• A versioning system will be in place to track model changes and ensure that updates do not disrupt the user experience.

CONCLUSION

The proposed methodology provides a robust, data-driven solution to bridging the gap between education and employment. By integrating data from educational institutions, job markets, and predictive analytics, the system will offer actionable insights, personalized recommendations, and real-time feedback to all stakeholders. This approach ensures that students are prepared for the future job market, educational institutions are aligned with industry needs, and employers are able to identify and hire the best talent.

Chapter 5

OBJECTIVES

The overarching goal of this project is to leverage advanced data analytics and machine learning to bridge the gap between education and employment. By providing detailed insights into the skills required in the job market and aligning them with educational outcomes, this project aims to enhance employability, facilitate informed decision-making by educational institutions, and streamline recruitment processes for employers. This chapter presents a detailed breakdown of the objectives, methodologies, and expected impacts of the project.

5.1 Enhancing Skills Alignment with Market Demands Through Data-Driven Insights

One of the key objectives of this project is to help educational institutions understand the constantly evolving demands of the job market. Educational programs are often slow to adapt to changes in industry requirements, resulting in a mismatch between the skills students acquire and the skills employers need. By analysing data from job postings, industry reports, and labour market statistics, this project will provide actionable insights that can guide curriculum development and program enhancements.

5.1.1 Job Market Trend Analysis

The first step in aligning education with market demands involves understanding the current trends in the job market. By using machine learning models to analyse a wide range of job-related data, including job postings, employer expectations, and the skills sought by different industries, the project will identify:

- Skills that are in high demand: By tracking keywords and the frequency with which certain skills are mentioned in job postings, we can identify both high-demand and emerging skills.
- Regional trends in employment: The project will also explore how employment trends vary by region, offering insights into which skills are most relevant in specific geographical locations.
- Industry-specific requirements: Tailoring insights to specific sectors (e.g., tech, finance, healthcare) will help educational institutions better understand the nuances of industry-specific skill needs.

The insights gained from this analysis will help identify skills that educational institutions should prioritize in their curricula to ensure students are better prepared for the workforce. This step is particularly important in a world where industries are rapidly changing due to technological advancements and globalization.

5.1.2 Curriculum Optimization Recommendations

Once we have a comprehensive understanding of market demands, the next objective is to provide recommendations for curriculum optimization. Based on the job market trends, the system will suggest modifications to existing programs or introduce new courses that address skills gaps. These suggestions will be:

- Actionable and data-driven: Using data from industry reports, job market trends, and skill analysis, we will recommend which subjects or skills should be included in the curriculum.
- Timely and adaptive: As market demands change, these recommendations will be updated in real-time, ensuring that educational institutions can remain agile and responsive to industry needs.

By helping institutions adapt their curricula, this project will contribute to closing the skills gap and enhancing the employability of graduates.

5.1.3 Impact on Educational Institutions

The effectiveness of this objective will be assessed based on how well educational institutions incorporate the suggested curriculum changes. Success metrics could include:

- Higher employment rates among graduates: As students are better equipped with market-relevant skills, their employability should improve.
- Industry feedback: Positive feedback from industry partners indicating that the graduates are well-prepared for roles.
- Curriculum updates and innovations: Tracking the number of changes made to educational programs based on these recommendations will also be a key measure.

5.2 Bridging the Skill Gap for Job Seekers with Personalized Career Guidance

A critical objective of this project is to provide job seekers with personalized career guidance to help them navigate the job market more effectively. By leveraging the data from both job seekers' profiles and the current state of the job market, this project will offer tailored career recommendations and skill development pathways that enhance employability.

5.2.1 Personalized Career Pathways

The system will utilize machine learning-based recommendation algorithms to suggest career paths for job seekers based on their educational background, work experience, and skill set. For example:

- Career path suggestions based on skills: If a user has experience in data analysis but lacks advanced programming skills, the system may recommend paths in data science or business intelligence, suggesting additional training in programming languages such as Python or R.
- Role matching: The system will match job seekers with potential job roles that fit their profiles, considering both qualifications and career aspirations.

These personalized career pathways will help job seekers make informed decisions about their future and pursue roles that align with both their capabilities and market demands.

5.2.2 Continuous Skill Development Recommendations

As the job market evolves, so too should the skill sets of job seekers. This objective ensures that the system provides ongoing guidance, recommending new skills, certifications, and courses that will help job seekers stay competitive. The system will:

- Monitor job market trends: By continually analysing job postings and industry needs, the system will identify new skills that are gaining importance and suggest relevant training.
- Provide lifelong learning opportunities: The system will emphasize the importance of continuous learning, suggesting online courses, workshops, and certifications that allow job seekers to upskill throughout their careers.

5.2.3 Impact on Job Seekers' Employability

The ultimate success of this objective will be measured by how many job seekers follow the career and skill development recommendations, as well as the rate at which these individuals secure employment. Key performance indicators (KPIs) include:

- Higher job placement rates: Tracking the success rate of users who follow the system's recommendations in securing jobs.
- Skill enhancement: Measuring the acquisition of new skills through the recommended courses or certifications.

5.3 Enhancing Employer-Job Seeker Matching Through Predictive Analytics

This objective focuses on improving the matching process between employers and job seekers by using predictive analytics. Predictive models will analyze both employer requirements and job seeker profiles to suggest optimal matches, streamlining the recruitment process.

5.3.1 Predictive Job Matching Models

Machine learning models will be developed to predict the most suitable candidates for job roles based on a variety of factors such as:

- Job seeker qualifications: Education, work experience, skills, and certifications.
- Employer preferences: Previous hiring data, job requirements, and role specifications.

The system will also take into account trends in job market demand, ensuring that the matches are not only based on static qualifications but are also future-oriented.

5.3.2 Real-Time Job Matching

This objective also aims to enable real-time matching, allowing employers to find suitable candidates quickly and efficiently. The system will:

- Use dynamic data: The job market and candidate pool will be constantly updated to provide real-time recommendations.
- Provide instant notifications: Employers will receive alerts when a new candidate that fits their requirements enters the pool.

Real-time matching will reduce the time and effort involved in recruitment, helping employers find the right candidates faster.

5.3.3 Impact on Hiring Efficiency

The success of this objective will be assessed based on:

- Reduction in hiring time: How much quicker employers can fill job vacancies with suitable candidates.
- Quality of hires: Tracking how well candidates perform in their roles once hired.

5.4 Automating Data-Driven Insights with Dashboards for Institutions and Employers

An essential part of this project is to empower both educational institutions and employers with data-driven dashboards that present actionable insights. These dashboards will allow users to visualize key trends, performance metrics, and predictions.

5.4.1 Dashboard for Educational Institutions

The dashboard for educational institutions will provide insights into:

- Skill gaps in curricula: Identifying areas where students lack relevant skills.
- Student employability data: Visualizing how well students from different programs are faring in the job market.
- Curriculum improvement suggestions: Highlighting areas where the curriculum can be updated to better prepare students for current job market demands.

5.4.2 Dashboard for Employers

For employers, the dashboard will present:

- Candidate pools: Information on available talent pools that match their job requirements.
- Industry trends: Insights into the skillsets that are becoming essential in the job market.
- Regional talent availability: Data that helps employers understand where the best talent for specific roles is located.

5.4.3 Impact on Decision-Making

The use of dashboards will enable both institutions and employers to make more data-driven decisions. This will be assessed by:

- The extent of dashboard usage: Tracking how frequently users interact with the dashboards and rely on the insights provided.
- Decision improvements: Measuring the improvement in hiring efficiency for employers and curriculum relevance for educational institutions.

5.5 Ensuring Data Privacy and Security for All Stakeholders

Given the sensitive nature of the data involved, ensuring robust data privacy and security is paramount. This objective focuses on implementing strong data protection measures to safeguard user information.

5.5.1 Data Encryption and Secure Storage

To protect personal data, the system will use industry-standard encryption techniques, including AES-256 encryption for storing sensitive data and SSL/TLS for secure data transfer. These measures will ensure that all user information is kept confidential and secure.

5.5.2 Role-Based Access Control

Role-based access control (RBAC) will ensure that only authorized users have access to sensitive information. This will be particularly important in ensuring that institutions, employers, and job seekers can only view data relevant to their role.

5.5.3 Compliance with Data Protection Regulations

The project will adhere to regulations such as GDPR and CCPA, ensuring that users' privacy rights are respected. Features such as data anonymization and the ability for users to delete or update their data will be implemented to comply with data protection laws.

Technology / Tool	Description / Usage	Importance in Project
Azure Data Lake Gen2	Cloud-based storage used to store raw and processed data.	Centralized and scalable data storage solution for efficient access and management.
Azure Databricks	Cloud-based Apache Spark environment for big data processing and analytics.	Enabled distributed data processing and transformation using PySpark.
PySpark	Python API for Apache Spark used for cleaning, transforming, and preparing large data.	Performed scalable and efficient data preprocessing for analysis and visualization.
Power BI	Business intelligence tool used to create dashboards and visual reports.	Delivered interactive visualizations to communicate insights and hiring trends.
DAX (Data Analysis Expressions)	Formula language in Power BI used to create calculated columns and custom metrics.	Enabled creation of meaningful KPIs and enhanced dashboard interactivity.
Data Mining (Manual/Web)	Collected data on TPOs and hiring companies across cities (name, email, platform, role).	Built the foundation dataset required for further analysis and visualization.
Data Cleaning	Standardization, deduplication, null handling, and format unification.	Ensured data accuracy and consistency before further processing and storage.

Table 5.5

Technical Components Overview

CONCLUSION

The objectives of this project aim to enhance the connection between education and employment by leveraging data analytics and machine learning technologies. By providing educational institutions with curriculum optimization insights, job seekers with personalized career guidance, and employers with predictive matching tools, the project promises to create a more efficient, transparent, and responsive ecosystem. Moreover, the focus on data security and privacy ensures that all stakeholders can trust the platform with their sensitive information, driving better decisions for all involved.

Chapter 6

SYSTEM DESIGN & IMPLEMENTATION

The System Design and Implementation chapter outlines the technical approach followed during the internship project, focusing on building a robust and scalable data analysis pipeline using cloud-based tools. The aim was to design a system capable of efficiently collecting, processing, analysing, and visualizing data related to the employment and educational landscape across India. This system supports strategic insights into bridging the gap between job market demands and educational training offered by engineering colleges.

6.1 System Architecture

The system was designed using a modular pipeline approach, leveraging cloud technologies and big data frameworks. It consisted of two key phases: data mining and data analysis, with a transition from local tools to cloud-native platforms to ensure scalability and performance.

6.1.1 Data Mining Phase

The initial phase focused on collecting large-scale structured and semi-structured data from various online sources:

- Sources: Engineering college TPO (Training and Placement Officer) contacts, company recruitment details, job role specifications, and skill requirements across domains.
- Tools Used: Microsoft Excel was used to organize and store the data initially.
- Approach: Manual and semi-automated techniques were used for extracting relevant records such as job descriptions, required skills, eligibility criteria, company profiles, and college training modules.

6.1.2 Data Analysis Phase

Once the raw data was collected, it was cleaned, structured, and analysed using powerful tools to uncover patterns and insights:

- Data Cleaning and Structuring:
    - Removal of duplicates and irrelevant data entries
    - Categorization of job roles by domain and skillsets
    - Mapping college training programs to market-required competencies
- Tools Used:

- Microsoft Excel & Power BI for early-stage visualization and exploratory analysis
- Azure Data Lake Storage (ADLS) for storing large datasets on the cloud
- Azure Databricks (ADB) powered by PySpark for distributed processing and in-depth analysis

---

6.2 System Components

Since no front-end or relational databases were involved, the system's components focused solely on data ingestion, processing, and visualization.

6.2.1 Azure Data Lake Storage (ADLS)

- ADLS was used as the central data repository for storing structured and unstructured data.
- Files from Excel were uploaded to ADLS in CSV/Parquet format.
- Its hierarchical namespace allowed better organization of college and company datasets.

6.2.2 Azure Databricks (ADB) with PySpark

- ADB enabled distributed data processing, which was ideal for large-scale education and employment datasets.
- PySpark scripts were used to:
  - Clean data (null values, inconsistencies)
  - Perform domain-wise grouping of job roles
  - Analyse skill trends across industry verticals
  - Identify gaps between college training programs and real-world requirements

6.2.3 Power BI for Visualization

- Power BI dashboards were created to visualize key metrics such as:
  - Most demanded job roles by sector (IT, core, etc.)
  - Skill frequency heatmaps
  - Regional distribution of company recruitments
  - Training offerings vs. market trends
- These visuals supported decision-making for stakeholders aiming to improve college-company alignment.

6.3 Data Flow and Integration

The project followed a streamlined data flow approach without complex integrations:

1. Data Collection:
   - Data gathered via Excel and structured manually.
2. Upload to Cloud:
   - Cleaned data uploaded to ADLS.
3. Processing:
   - PySpark used in ADB for performing transformations and computations.
4. Visualization:
   - Insights exported and fed into Power BI reports.

This modular pipeline ensured that each step—from collection to analysis—remained independent, scalable, and easily maintainable.

6.4 Security and Compliance

Even though the project didn't involve sensitive personal data, best practices were followed in terms of access control and cloud security:

- Azure Role-Based Access Control (RBAC) ensured that only authorized users could access ADLS and Databricks.
- Data Anonymization: Any college-specific identifiers were anonymized before visualization to preserve institutional privacy.
- Data Backup and Recovery: Data was versioned and backed up periodically on ADLS to prevent accidental loss.

6.4.1 Role-Based Access Control (RBAC) Implementation

Azure's RBAC policies were applied to enforce identity-based access. This ensured that only authorized team members could interact with sensitive storage and compute environments on ADLS and Databricks.

6.4.2 Audit Logging and Monitoring

Activity logs in Azure provided traceability for access and operations on data resources. This ensured accountability and supported future audits or evaluations.

6.4.3 Data Lifecycle Management

To maintain storage efficiency, lifecycle policies were defined for archiving or purging outdated datasets. This approach ensured that only relevant, recent data remained active in the pipeline.

6.5 Scalability and Performance Optimization

This section outlines how the system was designed for large-scale data workloads and optimized for future extensibility.

6.5.1 Distributed Architecture with PySpark

PySpark's parallel processing capabilities were crucial in handling datasets from hundreds of colleges and job portals across multiple states. Tasks like filtering, joining, and aggregating large tables were performed efficiently.

6.5.2 Incremental Processing

Instead of reprocessing the entire dataset on every run, the system allowed for incremental updates—only newly added data was ingested and transformed. This improved efficiency and reduced computational costs.

6.5.3 Partitioning and File Format Optimization

Data was stored in optimized formats such as Parquet, and partitioned by state and domain (e.g., IT, mechanical, etc.), which significantly improved read performance in both Databricks and Power BI.

6.6 Maintainability and Extensibility

Ensuring the system remains usable for future enhancements was a core design principle.

6.6.1 Modular Codebase

PySpark scripts were written in a modular, function-based structure—allowing developers to reuse and adapt logic for new datasets, such as different academic years or industries.

6.6.2 Config-Driven Workflows

Processing logic (e.g., file paths, filters, output schemas) was abstracted into configuration files. This allowed changes in parameters without modifying the core codebase.

6.6.3 Ease of Integration

Although the system didn't include a frontend, its architecture allows future integration with:

- APIs (for real-time job feed ingestion)
- Web portals (for TPO dashboards)
- ML models (for placement prediction and skills recommendation)

6.7 System Limitations

Despite its effectiveness, the system has certain limitations that constrain its current capabilities.

6.7.1 Manual Data Collection Bottlenecks

Initial data mining relied heavily on manual or semi-automated techniques. This introduces time delays, risk of human error, and scalability challenges when covering broader geographies or more institutions.

6.7.2 Static Data Snapshots

The current system processes data as static snapshots. Real-time data ingestion and continuous updates are not yet implemented, which limits responsiveness to dynamic changes in job markets.

6.7.3 Absence of Relational or Metadata Indexing

Without a structured database or metadata management system, data retrieval depends on file organization alone. This can hinder advanced querying or multi-parameter filtering at scale.

6.7.4 Limited Stakeholder Accessibility

While Power BI provides accessible visualizations, users without cloud credentials (e.g., non-technical college staff or recruiters) may face hurdles in directly accessing raw or intermediate data layers.

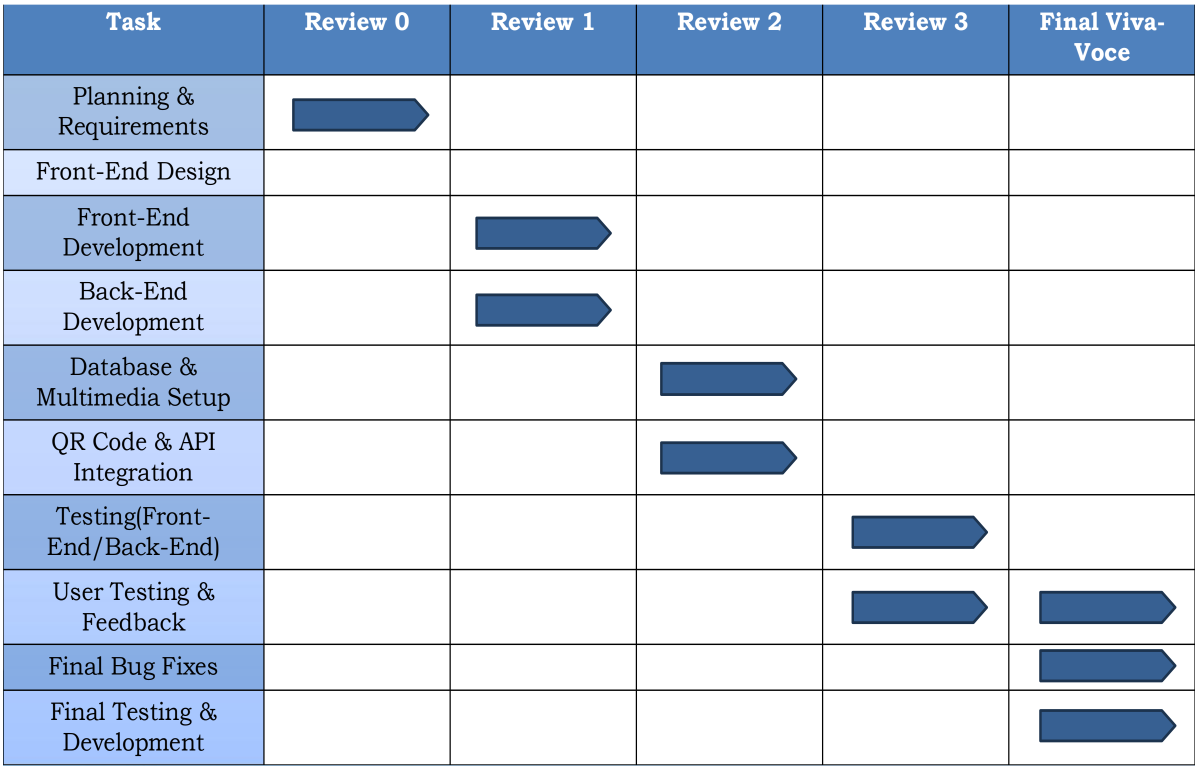
CONCLUSION

The system was designed to deliver a data-driven, scalable solution for analyzing the education-employment ecosystem. By leveraging cloud technologies like ADLS and ADB, combined with the flexibility of PySpark and the visualization power of Power BI, the system enabled comprehensive insights into job market alignment with educational training. The pipeline operated without the need for front-end interfaces or traditional databases, focusing solely on backend processing to maximize efficiency and clarity.

Chapter-7

TIMELINE FOR EXECUTION OF PROJECT

(GANTT CHART)



Chapter 8

OUTCOMES

The outcomes of this project reflect its potential to transform how educational institutions and industries align their efforts to close the skill-employment gap. By leveraging data analytics tools and cloud platforms, the project provided critical insights into training needs, market demand, and employability trends across engineering colleges and companies in India. The outcomes are multi-faceted, affecting stakeholders across the educational ecosystem, from academic policymakers to corporate recruiters and students.

8.1 Improved Understanding of Industry Demands

One of the most impactful outcomes was the detailed mapping of industry skill requirements based on job listings and recruitment data.

8.1.1 Skill Demand Heatmaps

Using PySpark and Power BI, we generated skill frequency heatmaps across various sectors—IT, core engineering, finance, and analytics. These visualizations highlighted which skills (e.g., Python, SQL, AutoCAD, Excel, ML) were in the highest demand and their distribution across companies and locations.

8.1.2 Job Role Clustering

Data analysis revealed clusters of job roles and their common required competencies. For instance, “Data Analyst” and “Business Analyst” roles often overlapped in required skills but differed in domain knowledge expectations. Such clustering helps colleges tailor specific modules to industry needs.

8.1.3 Domain-Wise Insights

We observed domain-specific hiring patterns—core mechanical jobs focused on tools like CATIA and SolidWorks, while IT jobs leaned heavily on programming and data tools. These findings offer actionable intelligence for curriculum refinement.

8.2 Institutional Assessment and Benchmarking

The project also produced valuable insights for educational institutions to assess their current training strategies.

8.2.1 TPO Data Consolidation

TPO (Training & Placement Officer) contact data was centralized across hundreds of colleges, making it easier for institutions to benchmark themselves in terms of placement trends, partner companies, and training initiatives.

8.2.2 Gap Analysis Between Training and Hiring

By comparing company job descriptions with college training modules, we identified critical mismatches—such as outdated software training or lack of project-based learning. This outcome encourages institutions to modernize their skill delivery models.

8.2.3 Inter-College Comparison Dashboards

Power BI dashboards enabled dynamic comparisons between colleges based on hiring stats, domain strengths, and training infrastructure. These insights help TPOs strategize collaborations and workshops.

8.3 Scalable and Cloud-Enabled Data Processing

The adoption of Azure Data Lake Storage and Azure Databricks ensured that the system could handle large datasets without performance degradation.

8.3.1 Efficient Big Data Handling

ADLS and PySpark enabled fast, distributed processing of employment and education datasets, improving runtime efficiency and reducing analysis complexity even as the dataset scaled across states and domains.

8.3.2 Modular and Reusable Pipelines

The data processing pipelines were modular, allowing reuse in future studies. Whether updating college training catalogues or job listings, the system remains robust without needing architectural redesign.

8.3.3 No-Code/Low-Code Accessibility

Many operations—especially dashboards and visualizations—were conducted using tools like Power BI and Excel, allowing non-technical stakeholders (e.g., TPOs or curriculum designers) to interact with insights without coding expertise.

8.4 Enhanced Decision-Making for Stakeholders

The outcomes of the project empower different categories of users to make better, data-backed decisions.

8.4.1 For Academic Institutions

- Align courses and workshops with real-world job roles.
- Invest in high-demand skill development (e.g., cloud tools, programming languages).
- Collaborate with companies actively recruiting in niche domains.

8.4.2 For Policy Makers

- Use data to fund educational reforms in underperforming regions or colleges.
- Design employment-linked skilling schemes based on real-time data insights.

8.4.3 For Students and Job Seekers

- Identify in-demand skills before graduation.
- Choose internships or certifications based on emerging industry needs.
- Get personalized recommendations based on data-driven job-role alignment.

8.5 Operational Efficiency and Future Readiness

Though the project does not use a traditional application interface, the backend infrastructure is optimized for future extensions.

8.5.1 Cloud-Native Scalability

The use of ADLS and ADB ensures that future datasets—such as new academic years or updated job roles—can be easily integrated without redesign.

8.5.2 Cost-Efficient Architecture

As no high-maintenance frontend or database layers were involved, the project remains cost-effective, relying on batch uploads, distributed processing, and dashboard-based reporting.

8.5.3 Foundation for Predictive Analytics

The current framework sets the stage for future integration of ML models to predict:

- Placement probability by college or course
- Demand trends for upcoming technologies
- Personalized training roadmaps for students

8.6 Data Democratization and Stakeholder Accessibility

The project emphasized usability and accessibility for a wide range of stakeholders, ensuring data insights are not limited to technical users.

8.6.1 Role-Based Dashboards

Customized Power BI dashboards were created to cater to different users—TPOs, academic deans, and hiring partners. These dashboards emphasized relevant KPIs, such as placement ratios, skill alignment scores, and regional hiring intensity.

8.6.2 Simplified Data Access

By utilizing cloud-based storage and visualization tools, stakeholders could access insights remotely without the need for high-end hardware or complex software installations.

8.6.3 Self-Service Analytics

With low-code interfaces and pre-configured filters, non-technical users could manipulate and explore data independently—enabling self-driven decision-making without requiring constant technical support.

8.7 Research Potential and Academic Impact

This project lays the groundwork for future research in the fields of education analytics, employment forecasting, and digital transformation in academia.

8.7.1 Dataset for Academic Research

The structured and cleaned datasets across TPOs, job roles, and hiring platforms provide a valuable resource for educational research, skill development studies, and labor market analysis.

8.7.2 Publishable Insights

Visualizations and analyses from this project can support publishable research papers in domains like EdTech, data science in education, and labor economics.

8.7.3 Cross-Institutional Collaboration

The project model encourages inter-college partnerships for shared training programs, benchmarking, and collaborative analytics initiatives, fostering a data-driven academic ecosystem.

CONCLUSION

The outcomes of this data-driven project establish a crucial link between education and employability by transforming raw datasets into meaningful insights. From identifying skill gaps and improving training alignment to offering strategic guidance to institutions, the system enables a more informed, responsive, and scalable approach to bridging India’s skill-employment divide. Powered by tools like PySpark, Power BI, and Azure Data Lake, the backend-focused architecture ensures both operational efficiency and extensibility for future research and policy development.

Chapter 9

RESULTS AND DISCUSSIONS

This project aimed to create a data-driven system to bridge the gap between education and employment by collecting, analysing, and interpreting data from engineering colleges and recruiters across India. The outcomes achieved through this project are multi-dimensional, contributing not only to academic knowledge but also offering practical value to educational institutions, students, and industries seeking talent. The project has led to improvements in data visibility, understanding job market trends, and helping educational stakeholders align their curricula with evolving industry requirements.

9.1 Structured Aggregation of Institutional Data

A key outcome of this project is the successful aggregation and organization of data related to engineering colleges and their Training and Placement Offices (TPOs). Prior to this, there was a lack of centralized access to such information, which limited the ability to understand institutional engagement with employers.

9.1.1 Enhanced Transparency Across Colleges

By collecting details such as contact information, student placement data, and recruiter engagement, the project has provided a more transparent view of how institutions are performing in terms of employability. This can encourage healthy competition among colleges to improve their placement support and collaboration with industries.

9.1.2 Institutional Benchmarking

With structured data in place, institutions can now benchmark their placement success rates against similar colleges. This comparative analysis supports administrative decision-making, helping colleges identify areas where they lag and emulate the practices of top-performing institutions.

9.1.3 Improved Discoverability for Recruiters

Recruiting firms often struggle to identify lesser-known but capable institutions. By making detailed TPO data available in a standardized format, this project opens new recruitment channels for companies beyond Tier-1 colleges, promoting inclusivity and wider access to opportunities.

9.2 Insightful Analysis of Job Market Demands

The project has made significant contributions in analysing and visualizing job market demand across various domains, using tools such as Power BI and PySpark.

9.2.1 Demand Mapping by Domain and Skills

Through data mining of job listings, the project mapped demand across sectors such as IT, core engineering, data science, and emerging technologies. This allowed clear identification of high-demand skills, including Python, cloud computing, AI/ML, and cybersecurity.

9.2.2 Regional and Temporal Trends

Another important outcome is the detection of geographic and temporal trends in job demand. By analysing data based on cities, states, and monthly hiring spikes, the project uncovers where and when certain skills are most in demand—useful for students planning their job search or higher studies.

9.2.3 Alignment with Curriculum Design

The insights from this analysis allow educational institutions to revisit their syllabus design. If, for instance, high demand is observed in DevOps and cloud computing, colleges can integrate these topics into elective courses or workshops. This increases the employability of graduates and strengthens industry-academia relevance.

9.3 Scalable Data Processing Using Cloud Tools

The shift from manual data handling to a cloud-based processing pipeline significantly improved the scalability and efficiency of the system. Microsoft Azure's Data Lake Storage (ADLS) and Azure Databricks (ADB) were used to process large volumes of semi-structured data.

9.3.1 Faster Processing and Better Resource Management

By moving to a PySpark-based approach, data was processed at scale in a distributed environment. This eliminated latency issues associated with local processing and enabled the handling of tens of thousands of rows without significant delays.

9.3.2 Real-Time Pipeline Possibilities

Although the current system is batch-oriented, the architecture supports future real-time updates. For instance, new job postings or updated TPO details could be automatically streamed into the database with minimal manual intervention.

9.3.3 Robust Error Handling and Logging

Through PySpark scripts and notebooks, errors in data extraction, cleaning, or transformation were logged, allowing easier debugging and greater system reliability—key outcomes for any data-centric system.

9.4 Practical Value to Stakeholders

The real-world utility of this project is one of its most impactful outcomes. Multiple stakeholder groups benefit from the findings and deliverables.

9.4.1 For Students

Students gain a clearer understanding of which skills to prioritize based on real market demand. They can also identify companies actively hiring in their domain, along with key hiring seasons, giving them a strategic edge in job applications.

9.4.2 For Colleges and Faculty

Colleges now have a blueprint to improve industry engagement. Faculty can introduce workshops or value-added programs in alignment with identified gaps in skill development. TPOs can also use the insights to approach relevant recruiters more effectively.

9.4.3 For Recruiters

Recruiters gain visibility into colleges with the right talent pool, even those in smaller towns or non-metro areas. This democratization of access reduces their hiring costs and opens new channels of talent acquisition.

9.5 Contribution to Research and Academia

As this project is being submitted for a conference and will be published as a research paper, it also makes valuable academic contributions.

9.5.1 Establishing a Methodological Framework

This project proposes a structured methodology for similar data-driven education-employment studies. From data mining to cloud-based transformation, the approach can be replicated and adapted for various sectors beyond engineering education.

9.5.2 Novel Integration of Tools

The integration of Excel, Power BI, ADLS, and ADB in a single pipeline showcases a cross-functional toolchain. This offers an end-to-end model for educational researchers looking to perform large-scale, impactful studies.

9.5.3 Future Scope for Comparative Studies

The current study lays the groundwork for future cross-country or multi-university comparative studies. Outcomes from this project can be expanded into longitudinal research to track changes in employability trends over time.

9.6 Limitations and Recommendations for Enhancement

While the outcomes are promising, there are some limitations that were encountered, which also provide direction for future work.

9.6.1 Data Availability and Consistency

Some TPOs or companies did not have updated or complete data online. Manual curation introduced occasional inconsistencies, suggesting a need for automated data collection pipelines (e.g., using web scraping or APIs) in future iterations.

9.6.2 Lack of Soft Skills and Internship Data

Although technical skills and job roles were well documented, the role of soft skills, internships, or extracurricular achievements could not be captured due to lack of structured data. This is a critical area for expansion.

9.6.3 Generalizability Across Domains

The system is tailored to engineering domains. Expanding it to include arts, commerce, or other professional degrees would enhance its scope and impact across educational landscapes.

9.7 Strategic Impact on Bridging the Skill Gap

At a macro level, this project helps create a foundation for reducing the industry-academia disconnect.

9.7.1 Evidence-Based Decision-Making

Educational institutions can now make data-backed decisions on curriculum updates, placement strategies, and training programs instead of relying on anecdotal feedback or assumptions.

9.7.2 Strengthening the National Talent Pipeline

When colleges align their training with actual market demands, graduates are more job-ready. This strengthens the overall talent pipeline for the country, particularly in fast-evolving tech sectors.

9.7.3 Enabling Policy Interventions

Insights from this study could also support policymakers in shaping national-level educational reforms or skilling initiatives like NEP 2020 or Skill India. A well-informed strategy is essential to prepare India’s youth for global competitiveness.

CONCLUSION

The outcomes of this project go beyond technical achievements—they reflect a step toward transforming how institutions and industries collaborate. By leveraging real data, applying scalable analytics tools, and delivering actionable insights, this project creates a bridge between education and employment that is both practical and replicable. The impact of this work is seen not only in better institutional planning and improved student employability but also in setting the stage for broader academic and industry innovations.

Chapter 10

CONCLUSION

The successful completion of this project marks a meaningful stride toward addressing one of the critical challenges in the academic and professional landscape—bridging the growing gap between educational output and employment demands. By adopting a data-centric methodology powered by modern cloud technologies, this project presents a practical and scalable framework to assess, understand, and align institutional efforts with industry requirements. Through the integration of Microsoft Azure's Data Lake Storage (ADLS), Azure Databricks (ADB), and PySpark along with traditional tools like Excel and Power BI, the project offers an effective solution for handling and analysing large-scale datasets that reflect the real-world employment ecosystem.

This system was developed with a dual objective: first, to provide insights into how well institutions are preparing students for the workforce, and second, to guide stakeholders in making evidence-based decisions to enhance employability. The initial phase involved data mining from a wide network of engineering colleges and recruiters across India, particularly focusing on the activities and contact points of Training and Placement Officers (TPOs) and the hiring trends of companies. In the next phase, this raw data was transformed, cleaned, and analysed to derive actionable insights that could be visualized and interpreted effectively.

One of the most notable achievements of this work is its transformation from a conventional, static approach to a cloud-based, automated, and scalable architecture. Data previously collected in fragmented forms was systematically organized and analysed using PySpark in the Databricks environment. This ensured that processing could be done efficiently, even for large and complex datasets, while maintaining flexibility for future scalability. The use of Power BI further enabled the development of interactive dashboards that present trends in an intuitive, user-friendly manner, making the findings accessible to both technical and non-technical stakeholders.

The insights derived from this system highlight the nature and depth of the disconnect between educational institutions and employers. It was observed that despite the availability of large numbers of graduates, many companies struggle to find candidates with the right skill sets. Conversely, students often lack a clear understanding of the actual requirements of the job market, resulting in missed opportunities and unproductive placement efforts. Through this system, it became evident that emerging sectors like Artificial Intelligence, Data Science, and

Cloud Computing are experiencing higher demand, whereas traditional fields like embedded systems and legacy software roles are showing signs of stagnation. These insights are not only valuable to colleges looking to revise curricula but also to students aiming to make strategic choices in terms of electives, certifications, and internships.

Furthermore, the project uncovered temporal trends in hiring cycles, geographic preferences in company recruitment patterns, and the types of roles companies are offering to students. By clustering companies based on their hiring profiles and matching them with student aspirations and capabilities, a meaningful correlation could be established. This can assist institutions in building better industry relationships and preparing students for the most relevant opportunities. The project has also shown that such data-driven frameworks can guide policy makers and curriculum designers in shaping a more adaptive and responsive education system.

The use of Microsoft Azure and its suite of tools significantly contributed to the project's flexibility and robustness. Azure Data Lake Storage facilitated the secure and scalable storage of semi-structured and unstructured data. Azure Databricks, with its integration of Apache Spark, enabled high-performance distributed data processing, while PySpark scripts made the transformation pipeline efficient and reusable. These tools worked cohesively to support advanced data analytics without the need for a front-end or database, aligning with the project's goal of remaining backend-heavy but technically sound and cloud-native.

Despite the success of this system, certain limitations remain. The quality of collected data posed challenges due to inconsistencies, outdated information, and lack of standardization among institutional websites. Many TPO pages lacked current placement data, while job listings on company websites often varied in terminology and structure. This required significant effort in manual cleaning and formatting. Another limitation was the exclusion of qualitative data—such as student portfolios, soft skills, and internship experience—which are increasingly important in hiring decisions. Future enhancements could incorporate this kind of data for a more holistic view.

Additionally, while this project offers a strong foundation, its impact will grow substantially when longitudinal tracking is introduced. Observing how placements evolve over multiple academic years will allow the system to detect deeper patterns and provide strategic forecasting. This capability would empower institutions to take timely corrective actions, and enable students to make more forward-looking decisions regarding their career paths. Real-time feedback mechanisms, recruiter rating systems, and student performance analytics could be added to further enrich the model.

Looking forward, the system could be expanded to include predictive analytics, enabling it to forecast job market shifts and recommend skill development pathways for students based on emerging industry trends. A future version of the platform could also include automated alerts for institutions and students about in-demand technologies, upcoming recruitment drives, and industry certification programs. This would make the system more dynamic and responsive to change. Moreover, integration with national employment databases or career platforms could give it national-level applicability, driving better coordination between academia and industry at a policy level.

In conclusion, this project delivers more than a research model—it offers a strategic blueprint that educational institutions, companies, and even governments can leverage to build stronger, more meaningful bridges between education and employment. It demonstrates that data, when thoughtfully collected, organized, and analyzed, can illuminate the root causes of systemic issues and point toward effective solutions. The tools and techniques applied in this project not only meet current technological standards but also allow for future adaptability and innovation.

CONCLUSION

This project addresses the critical disconnect between engineering education and employment readiness in India. By leveraging cloud-based technologies like Microsoft Azure (ADLS, ADB) and PySpark, alongside tools like Excel and Power BI, a scalable system was developed to mine, clean, and analyze placement-related data from colleges and companies.

The system provides actionable insights into job trends, skill gaps, and hiring patterns, helping institutions align their training with real-world industry needs. It highlights in-demand fields such as AI, Data Science, and Cloud Computing, while exposing the declining relevance of some traditional roles. Additionally, it helps map geographic and temporal hiring trends, improving institutional strategy and student preparedness.

While data inconsistencies and the exclusion of qualitative factors posed challenges, the system lays a solid foundation for future improvements such as predictive analytics, real-time alerts, and policy integration. Overall, this work is a step toward transforming India’s education-to-employment pipeline through data-driven decision-making and strategic alignment with industry demand.

REFERENCES

[1] Barton, D., Farrell, D., & Mourshed, M. (2012). "Education to Employment: Designing a System that Works." McKinsey & Company

[2] Zhu, G., Kopalle, N. A., Wang, Y., Liu, X., Jona, K., & Börner, K. (2020). "Community-Based Data Integration of Course and Job Data in Support of Personalized Career-Education Recommendations."

[3] Fernandes, J. (2023). "The Role of Data-Driven Decision-Making in Effective Educational Leadership." Academy of Educational Leadership Journal, 27(S2), 1-3

[4] Bhatia, M. K., Nigam, S., Adwani, S., & Kothari, D. (2024). "Applying Data-Driven Decision-Making to Academic Hiring Procedures: An HR Analytics in Higher Education." Library Progress International, 44(3).

[5] Sharma, R. (2024). "Boosting Student Employability with the Effective Use of Data." Hurix Digital.

[6] Martinez, K. (2021). "Data-Driven Decision-Making in Education: Using Data to Improve Instruction." Journal of Arts, Society, and Education Studies.

[7] Liu, Y. (2024). "Data-Driven Decision Making in Higher Education Institutions: State-of-play."

[8] BibliU. (2023). "Data-Driven Decision-Making in Education: A Guide for Higher Learning Leaders."

[9] McKinsey & Company. (2018). "Creating an Effective Workforce System for the New Economy."

[10] McKinsey & Company. (2020). "Closing the Skills Gap: Creating Workforce-Development Programs That Work for Everyone."

[11] McKinsey & Company. (2021). "Creating Opportunities for Meaningful Employment."

[12] National Association of Colleges and Employers. (2023). "Career Services Benchmark Survey."

[13] World Economic Forum. (2023). "The Future of Jobs Report 2023." Retrieved from

[14] OECD. (2022). "Education at a Glance 2022: OECD Indicators." OECD Publishing.

[15] UNESCO. (2021). "Global Education Monitoring Report 2021: Non-State Actors in Education."

APPENDIX-A

PSUEDOCODE

Source Code:
https://github.com/dhanushcrm/Data-Analytics

APPENDIX-B

SCREENSHOTS

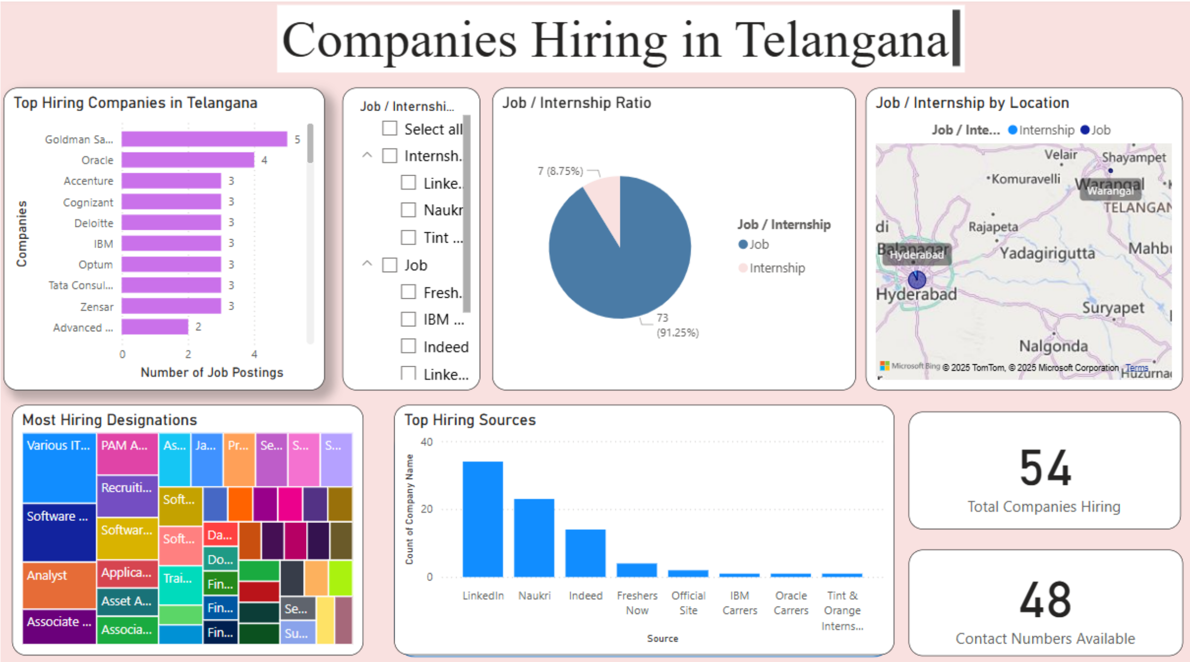
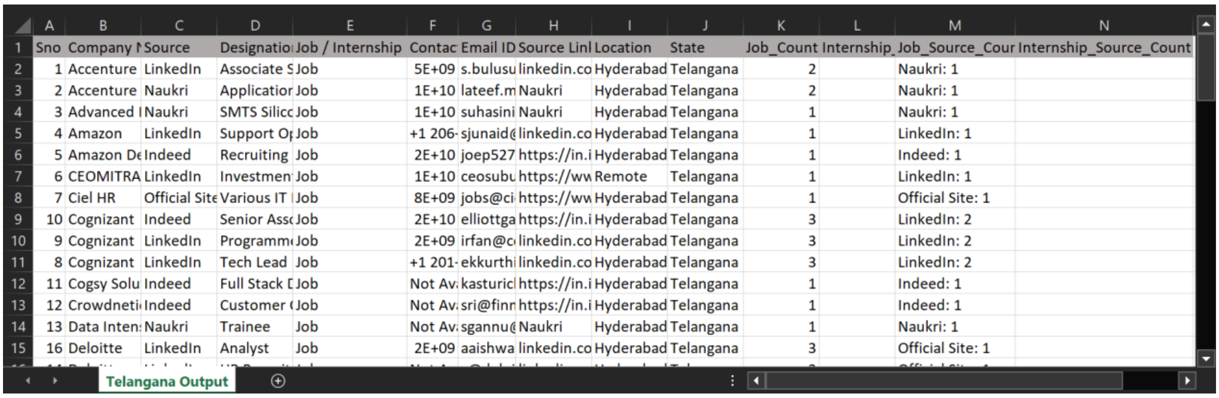


Figure B1

Visualization of Job Trends



Sno	Company N	Source	Designatio	Job / Internship	Contac	Email ID	Source Linl	Location	State	Job_Count	Internship_	Job_Source_Cour	Internship_Source_Count
1	Accenture	LinkedIn	Associate S	Job	5E+09	s.bulusu	linkedin.co	Hyderabad	Telangana	2		Naukri: 1	
2	Accenture	Naukri	Applicatior	Job	1E+10	lateef.m	Naukri	Hyderabad	Telangana	2		Naukri: 1	
3	Advanced I	Naukri	SMTS Silicc	Job	1E+10	suhasini	Naukri	Hyderabad	Telangana	1		Naukri: 1	
4	Amazon	LinkedIn	Support Op	Job	+1 206-	sjunaid@	linkedin.co	Hyderabad	Telangana	1		LinkedIn: 1	
5	Amazon De	Indeed	Recruiting	Job	2E+10	joep527	https://in.i	Hyderabad	Telangana	1		Indeed: 1	
6	CEOMITRA	LinkedIn	Investmen	Job	1E+10	ceosubu	https://ww	Remote	Telangana	1		LinkedIn: 1	
7	Ciel HR	Official Site	Various IT	Job	8E+09	jobs@ci	https://ww	Hyderabad	Telangana	1		Official Site: 1	
10	Cognizant	Indeed	Senior Ass	Job	2E+10	elliottga	https://in.i	Hyderabad	Telangana	3		LinkedIn: 2	
9	Cognizant	LinkedIn	Programm	Job	2E+09	irfan@c	linkedin.co	Hyderabad	Telangana	3		LinkedIn: 2	
8	Cognizant	LinkedIn	Tech Lead	Job	+1 201-	ekkurthi	linkedin.co	Hyderabad	Telangana	3		LinkedIn: 2	
11	Cogsy Solu	Indeed	Full Stack D	Job	Not Av	kasturicl	https://in.i	Hyderabad	Telangana	1		Indeed: 1	
12	Crowdneti	Indeed	Customer	Job	Not Av	sri@finr	https://in.i	Hyderabad	Telangana	1		Indeed: 1	
13	Data Inten	Naukri	Trainee	Job	Not Av	sgannu@	Naukri	Hyderabad	Telangana	1		Naukri: 1	
16	Deloitte	LinkedIn	Analyst	Job	2E+09	aaishwa	linkedin.co	Hyderabad	Telangana	3		Official Site: 1	

Telangana Output

Figure B2

Processed Data for performing EDA and Visualization

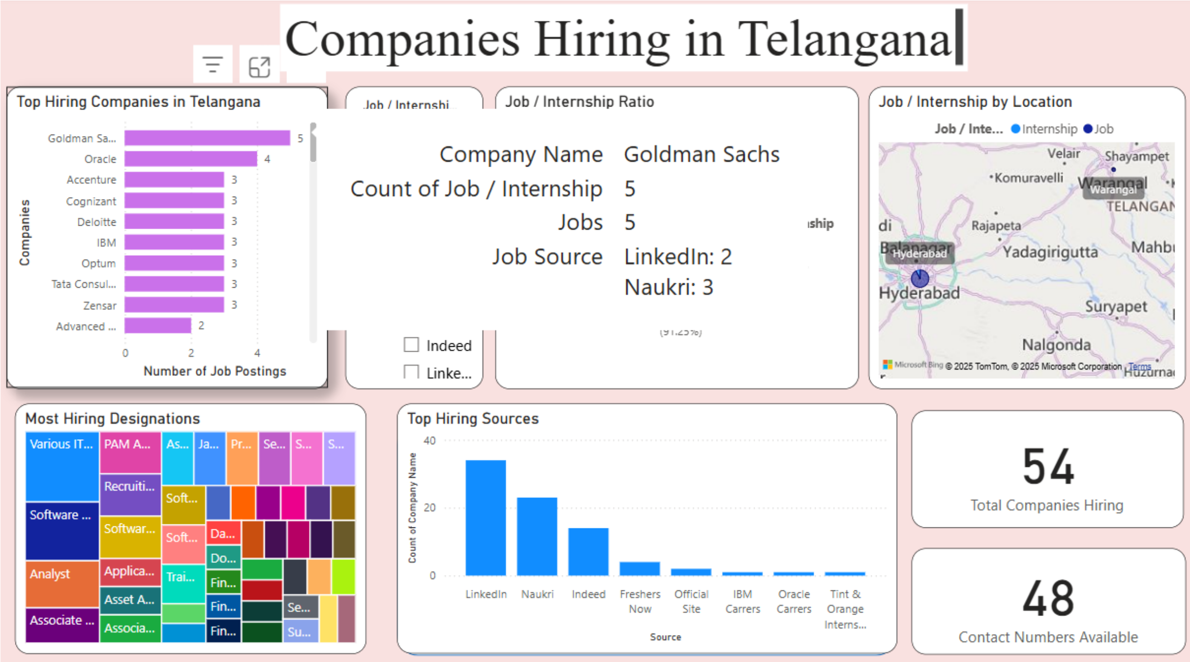


Figure B3

Highlighting the key metrics of Visualization

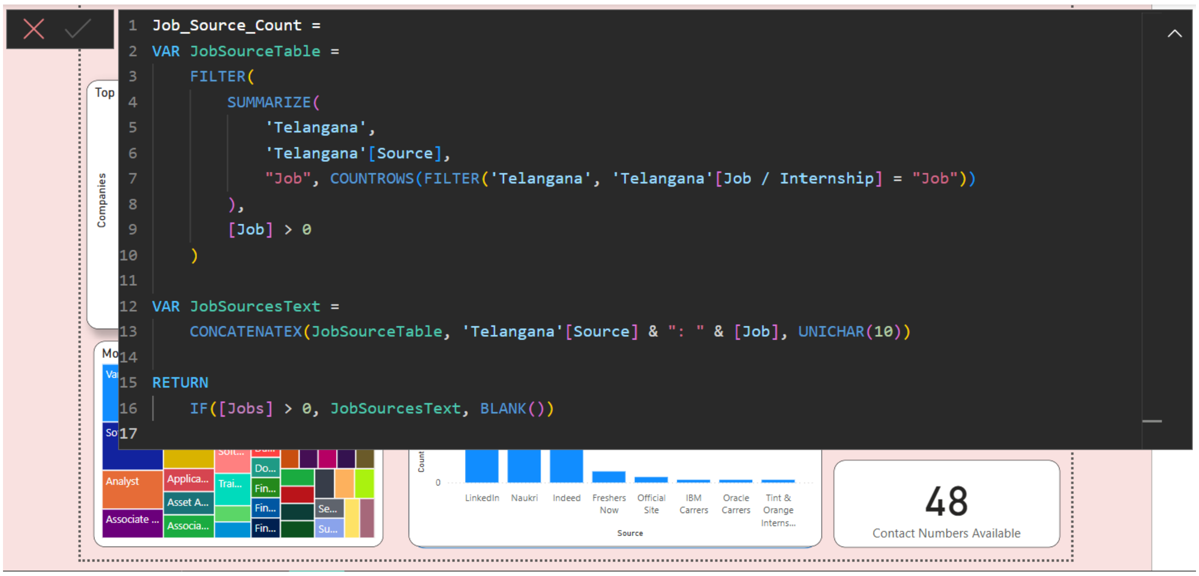


Figure B4

Performing DAX Query on the Dataset

APPENDIX-C

ENCLOSURES

PLAGARISM REPORT

turnitin Page 2 of 71 - Integrity Overview Submission ID trn:oid:::1:3245612105

10% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography

Match Groups

- 88 Not Cited or Quoted 10%
  Matches with neither in-text citation nor quotation marks
- 0 Missing Quotations 0%
  Matches that are still very similar to source material
- 0 Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- 0 Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

7%	Internet sources
6%	Publications
7%	Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

turnitin Page 2 of 71 - Integrity Overview Submission ID trn:oid:::1:3245612105

SUSTAINABLE DEVELOPMENT GOALS

1. SDG 4: Quality Education

Goal: Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all.

Target 4.5

By 2030, eliminate gender disparities in education and ensure equal access to all levels of education and vocational training for the vulnerable, including persons with disabilities, indigenous peoples, and children in vulnerable situations.

Project Alignment

Our platform's personalized learning paths and multilingual support ensure that learners from diverse backgrounds, including those with disabilities and non-native language speakers, have equitable access to educational resources and career guidance. By tailoring content to individual needs and preferences, the project promotes inclusive education and supports lifelong learning opportunities for all.

---

2. SDG 8: Decent Work and Economic Growth

Goal: Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all.

Target 8.6

By 2020, substantially reduce the proportion of youth not in employment, education or training.

Project Alignment

By analysing educational backgrounds and employment trends, the system identifies skill gaps and recommends targeted training programs. This proactive approach facilitates smoother transitions from education to employment, aiming to reduce youth unemployment and underemployment rates.

---

3. SDG 9: Industry, Innovation, and Infrastructure

Goal: Build resilient infrastructure, promote inclusive and sustainable industrialization, and foster innovation.

Target 9.5

Enhance scientific research, upgrade the technological capabilities of industrial sectors in all countries, particularly developing countries, including, by 2030, encouraging innovation and substantially increasing the number of research and development workers per 1 million people and public and private research and development spending.

Project Alignment

The integration of AI-driven analytics and machine learning models in the platform exemplifies technological innovation in the education and employment sectors. By leveraging data to inform decision-making, the project contributes to upgrading technological capabilities and encourages further research and development in these fields.

---

4. SDG 10: Reduced Inequalities

Goal: Reduce inequality within and among countries.

Target 10.2

By 2030, empower and promote the social, economic and political inclusion of all, irrespective of age, sex, disability, race, ethnicity, origin, religion or economic or other status.

Project Alignment

The platform's focus on personalized learning paths and employment opportunities ensures that individuals from marginalized communities receive support tailored to their unique circumstances. By addressing barriers to education and employment, the project fosters greater social and economic inclusion.

5. SDG 16: Peace, Justice, and Strong Institutions

Goal: Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels.

Target 16.10

Ensure public access to information and protect fundamental freedoms, in accordance with national legislation and international agreements.

Project Alignment

By providing transparent access to educational and employment data, the platform empowers users to make informed decisions. This transparency supports accountability and inclusivity in institutional processes related to education and workforce development.