Smart College Placement Portal Using Web Scraping and Sentiment Analysis

Ashutosh Kumar, Ankush Kumar, Monish Basaniwal, Nanditha C P

Abstract

The rapid expansion of online education and the increasing demand for skilled professionals have emphasized the need for effective college placement management portals. This abstract presents a comprehensive overview of an innovative online platform designed to streamline and optimize the college placement process.

Our online college placement management portal leverages advanced technologies to provide a user-friendly and efficient interface for students and educational institutions. The portal facilitates seamless communication and collaboration among all stakeholders, enhancing the overall effectiveness of the placement process.

For students, the portal offers a centralized platform to explore and apply for job opportunities posted by TPOs. It provides comprehensive profiles of participating companies, enabling students to make informed decisions about their career paths.

Additionally, the portal features personalized recommendations based on students' skills, interests, and academic performance, increasing the likelihood of successful placements.

The online college placement management portal also incorporates robust administrative features, allowing administrators to configure system settings, manage user accounts, and generate insightful reports. It ensures data security and privacy by implementing stringent authentication and encryption measures.

Overall, our online college placement management portal revolutionizes the college placement process by leveraging technology to create a seamless, efficient, and transparent ecosystem. By connecting students and educational institutions on a single platform, it fosters mutually beneficial collaborations and enhances the prospects of successful placements. With its user-friendly interface and powerful features, the portal empowers stakeholders to navigate the complex job market with confidence and ease.

Introduction

1.1 OVERVIEW

A college placement software is a computer program that helps colleges and universities manage the process of placing students into courses and academic programs. The software typically includes features such as:

- **Student data management**: The software stores and organizes student data such as

  transcripts, test scores, and builds a personalized resume.

- **Reporting and analytics**: The software can generate reports and analytics on student performance, course enrolment, eligibility for the drives, and other data on placements,

  to help colleges make informed decisions about their academic programs.

- **Mobile accessibility**: Many college placement software programs are designed to be

  accessible from mobile devices,

  allowing students and staff to access the system from anywhere.

Overall, college placement software helps colleges and universities streamline the process of

placing students into companies, and provides support for students as they plan their professional career.

1.2 OBJECTIVES

1. Streamlining the Placement Process: The primary objective of the software is to streamline the college placement process by automating manual tasks and reducing administrative burdens. It aims to create an organized and efficient system that simplifies the entire placement lifecycle.

2. Providing a Centralized Platform: The software aims to create a centralized platform where all

   placement-related information, resources, and activities are consolidated. This centralization simplifies access to information for students, TPOs, and college

   administrators, reducing redundancy and improving overall efficiency.

3. Personalizing the Placement Experience: The software focuses on personalizing the

   placement experience for students

   by providing tailored recommendations, insights, and guidance based on their individual profiles, skills, and preferences. It aims to match students with suitable

job opportunities and facilitate their career growth.

4. Enabling Data-Driven Decision Making: The software aims to

   collect and analyse

   placement data to provide valuable insights to educational institutions. It assists in

   identifying trends, evaluating the

   effectiveness of placement strategies, and making data-driven decisions to enhance the overall placement outcomes.

5. Ensuring Security and Privacy: The software prioritizes the security and privacy of user data by

   implementing robust authentication, encryption, and data protection measures. It ensures that

   sensitive information is safeguarded, building trust among users and maintaining compliance with relevant data protection regulations.

1.3 PROBLEM STATEMENT

Campus placements are conducted in all colleges for all the education fields. Various software and other sector companies are conducting campus recruitment process for selecting candidates. When campus selections are conducted the students have to provide their resume to the concern TPO officer for attending the campus interviews. This routine process is maintained manually. Many colleges have automated system but has some limitations. Students uploaded their CVs early, leaving them as it is in time. Lists were produced for each company based on the present information. Students have to check their mail regularly. Searching is done manually based on the company criteria and requirements.

The primary motto was to develop such a system which will reduce the load of the manual placement process and provide an effective way to conduct campus hiring. The main objective was to develop a platform where all the placements related activities can happen without any tedious manual work of dealing with large amounts of data.

1.4 PROPOSED SYSTEM

- **Student data management**: The system should store and organize student data such as transcripts, test scores, and be able to build resume.

- **Drive scheduling**: The system should help colleges create and manage drive schedules, including assigning students to it.

- **Advisor support**: The system should provide advisors with tools

to help students choose path and plan their professional careers.

- **Student self-servic**e: The system should include a portal that allows students to view their resume, request changes, and view their academic progress.

- **Communication tools**: The system should include tools for communication between students, TPOs, and Admin.

- **Reporting and analytics**: The system should generate reports and analytics on student performance, course enrolment, and other data on placement happening, to help colleges make informed decisions about their academic programs.

- **Integration with other systems**: The system should integrate with other systems used by the school, such as student information systems, financial aid systems, and learning management systems.

- **Mobile accessibility**: The system should be accessible from mobile devices, allowing students and staff to access the system from anywhere.

- **Customization**: The system should offer customization options to allow schools to tailor the software to their specific needs and processes. This can include custom fields, workflows, and user roles.

- **Data security**: The system should have robust security measures in place to protect student data and ensure compliance with relevant privacy regulations.

- **Scalability**: The system should be able to handle a large number of students and courses, and should be able to adapt as the organization's needs change over time.

- **User-friendly interfac**e: The system should have a user-friendly interface that is intuitive and easy to navigate, to reduce the learning curve for users.

- **Automation**: The system should include automation features to help streamline processes and reduce the workload for staff and students.

- **Integration with external systems**: The system should be able to integrate with external systems such as testing platforms and job boards, to provide students with additional resources and support.

- **Customized reporting**: The system should allow users to create customized reports to meet the specific needs of their organization.

- **Training and support**: The system should provide training and support

to help users get the most out of the software and ensure its effective use.

- **Data backup and recovery**: The system should regularly back up data to prevent loss in case of system failures or data corruption. It should also have mechanisms in place to recover data quickly and efficiently.

- **Document management**: The system should provide a centralized document repository where users can upload, store, and access relevant documents such as resumes, cover letters, and certifications.

- **Application tracking**: The system should enable tracking and monitoring of the application status for various job opportunities, providing students and TPOs with real-time updates on the progress of applications.

- **Feedback and evaluation**: The system should allow students and employers to provide feedback and

  evaluationsoninterviews,

  placements, and overall experiences, enabling continuous improvement of the placement process.

- **Data analytics and visualization**: The system should offer advanced data analytics capabilities and visualization tools to generate insights and trends from placement data, aiding decision-making and strategic planning.

- **Data privacy and consent management**: The system should comply with privacy regulations and provide features for managing user consent for data usage and sharing. It should also allow users to update their privacy settings and preferences.

- **Performance tracking**: The system should track and monitor the performance of students during the

  placement process, including metrics such as interview success rates, offer acceptance rates, and overall placement outcomes.

- **Data migration**: The system should have provisions for seamless data migration from existing systems to ensure a smooth transition and minimize data loss during implementation.

System Requirements

2.1 FUNCTIONAL REQUIREMENTS

- Students will be able to login on the platform and create a profile with their academic, personal, internship, certification, achievement, project details and also update their skillsets.

- Students will be able to see a list of available placement drives and view details such as the company profile, eligibility criteria, and application deadline. They will also be able to see why their profile is eligible/not eligible for that drive. Students will also receive timely intimations if their profile is matching the criteria set on a specific drive.

- Students will be able to apply for placement drives by submitting their resumes and other required documents. They will only be able to apply to the drives they are eligible for and track the application.

- The platform will automatically calculate student eligibility for each drive based on the criteria provided by the TPO and match it with the student's profile and shortlist the students based on it and send them a notification saying they are shortlisted for the drive.

- TPOs will be able to upload placement drive details on the platform, including the company profile, eligibility criteria, application deadline, and selection process.

- TPOs will be able to whitelist new companies onto the platform directly from Glassdoor, there will be no need of manual input from the TPO, search results are fetched directly from Glassdoor's official website.

- TPOs will be able to view and manage the list of students who have applied for each drive. They will be able to download the excel sheet.

- TPOs will be able to download reports and analytics on placement performance, such as the number of students placed, the average salary offered, and the top recruiting companies.

- Faculties will be able to view and manage the list of eligible students for each drive, and provide.

- Recommendations for students who they believe are well-suited for the drive.

2.2 NON - FUNCTIONAL REQUIREMENTS

- The platform will be secure and protect the personal and sensitive information of students, TPOs, and faculties. The entire authentication system is based on tokens which are generated dynamically at the time of login. Each token is tied to a specific user role and hence no user can perform any task which is allowed for that specific user only.

- The platform will be user-friendly and easy to navigate, with a clean and intuitive design. Each component has been designed with ease of use in mind with little to no learning curve attached with the platform.

- The platform will be scalable and able to handle a large number of users and drives without any downtime or performance issues. The platform is built on Node.js which is proven to handle a large number of requests in a single second. The platform also makes use of AWS Elastic Beanstalk to allow incoming requests to be balanced on the server between different deployments and auto-scale as required.

- The platform will be accessible from any device with an internet connection, including smartphones, tablets, and computers. The smartphone will have its own standalone app for Android and iOS separately mainly for the students to allow them to easily have a bird's eye view of the entire system and easily navigate, access and view their progress during the placement season.

- The platform will provide regular updates and maintenance to ensure that it remains functional and secure.

- The platform should have fast response times and low latency to ensure a smooth and efficient user experience. It should be able to

  handle simultaneous user interactions without significant delays.

- The platform should be highly reliable, minimizing system failures and downtime. It should have backup mechanisms and disaster recovery plans in place to ensure data integrity and availability.

- The platform should be compatible with different web browsers, operating systems, and devices, ensuring a consistent experience

across various platforms. It should adhere to web standards and accessibility guidelines.

2.3 SYSTEM REQUIREMENTS

Hardware Requirements

- Processor: Dual core/i3 and above
- RAM: 2GB and above
- Speed: 500MHz and above
- Secondary device: 250GB and above

Software Requirements

The whole platform is aimed to be extremely robust and dynamic with most of the components residing in a cloud environment.

- Base Programming Languages: JavaScript, Bash, Python, YAML 2 Language Frameworks: NodeJS, ReactJS, ExpressJS
- Database: MongoDB (Driver: mongoose)
- Caching Service: Redis
- Error Logging and Monitoring: Sentry/Rollbar
- Deployment: AWS Codedeploy
- Deployment machine provider: AWS EC2
- Load Balancing: Elastic Beanstalk + Elastic IPs
- CI/CD: AWS CodePipeline
- Object Storage: AWS S3
- Mailers: AWS SNS/Nodemailer
- Static Serving Provider: Vercel/Netlify

System Analysis

3.1 Block Diagram

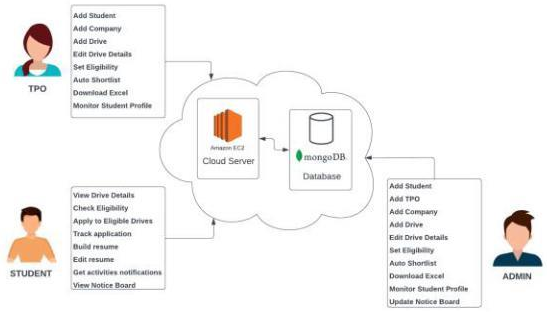

Figure 3.1: Block diagram for the overall system

3.2 Methodology

The online platform extensively utilizes Python and BeautifulSoup for web scraping purposes, specifically to extract company lists, reviews, and potential interview questions directly from Glassdoor. Subsequently, the collected reviews undergo sentiment analysis using Linear Regression Natural Language Processing algorithms. This analysis assigns sentiment scores to determine the positivity or negativity of each review. Ultimately, a cumulative score is calculated for each company, bounded by a maximum and minimum value.

To ensure system reliability and prevent a single point of failure, the scraping and analysis server has been separated from the primary modules. This segregation allows for appropriate restoration timelines in case of any disruptions.

In order to achieve a highly flexible, rapid, and stable system, the platform incorporates advanced and complex concepts from Database Management Systems (DBMS). These include Query Scheduling, Transactions, Crons, and sophisticated query-based matching and building techniques. Such implementations enable efficient handling of data, quick response times, and enhance system stability.

The entire platform is supported by the robust infrastructure and security provided by Amazon Web Services (AWS). It leverages various AWS services, such as EC2 for computing resources, Elastic IPs for flexible networking, Elastic Beanstalk Load Balancers for efficient traffic distribution, S3 storage for secure data storage, and SNS and SES for effective communication. Additionally, appropriate logging and monitoring policies are in place to ensure comprehensive oversight of system operations.

The online platform also utilizes Python's multiprocessing capabilities to parallelize the web scraping process, allowing for faster data collection from multiple sources simultaneously. In addition to sentiment analysis, the platform employs natural

language processing techniques like named entity recognition and topic modelling to extract relevant information from the collected reviews, such as key company attributes and trending topics.

To ensure data integrity and minimize duplicate entries, the platform employs data deduplication techniques during the web scraping process. This prevents redundant information from being stored and improves the overall quality of the extracted data. The platform employs data caching mechanisms to reduce the load on the database and improve response times. Frequently accessed data, such as company lists and popular interview questions, are stored in memory or in a separate caching layer to avoid the need for repeated database queries.

Advanced data visualization techniques, such as interactive charts and graphs, are implemented to present the collected information in a user-friendly and easily understandable format. Users can explore and analyse the data visually, gaining valuable insights at a glance.

Design

4.1 System Architecture

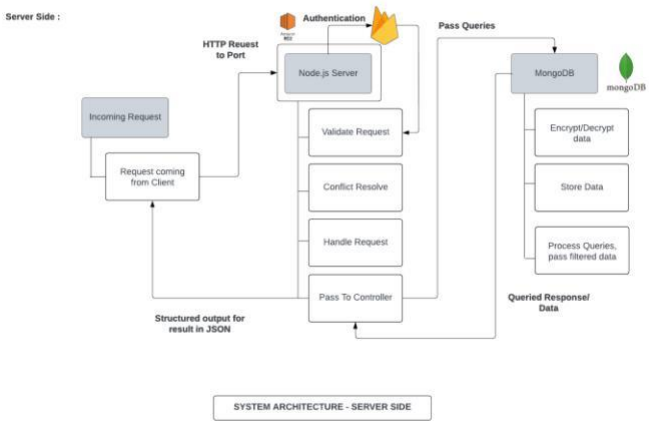

Figure 4.1.1: System Architecture design for Backend

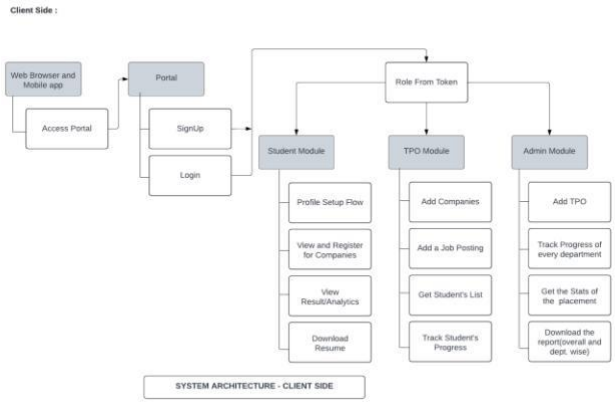

Figure 4.1.2: System Architecture design for Frontend

4.2 Flow Chart

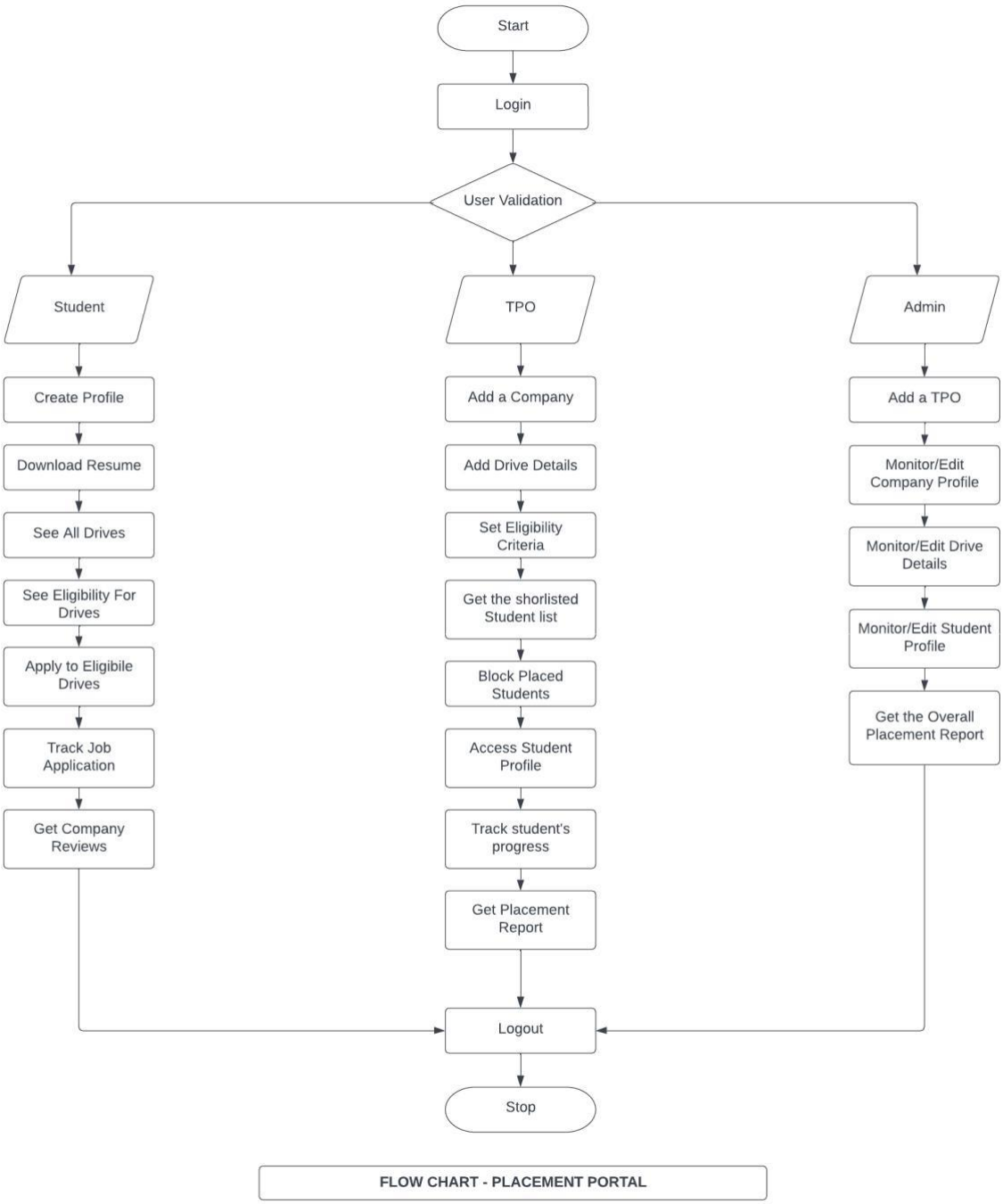

Figure 4.2: Flow Chart for the overall system

4.3 Data Flow Diagram

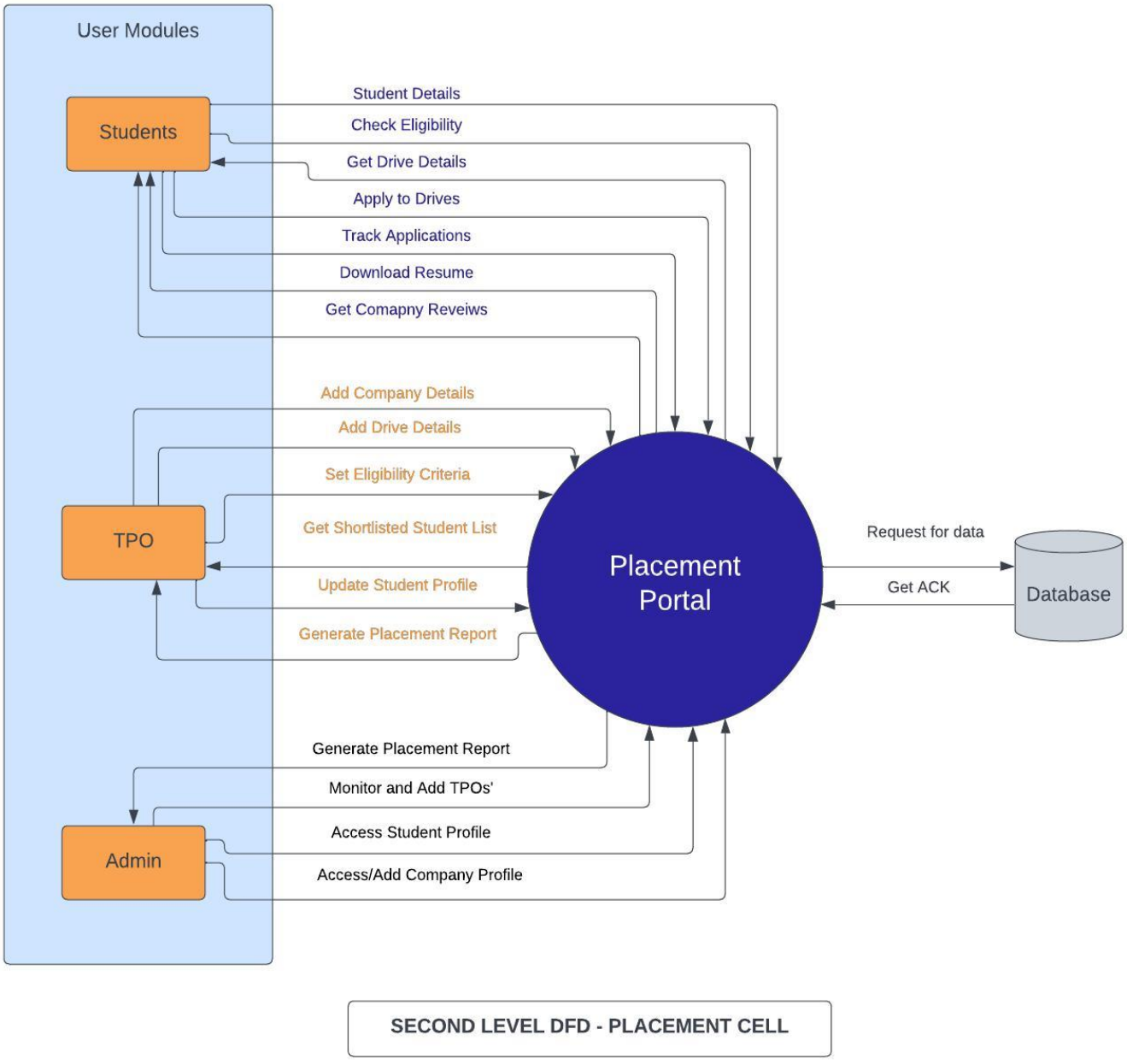

Figure 4.3: Second Level Data Flow Diagram for The System

Implementation

Student Module:

- The student module is designed for students to access and apply for job opportunities offered various companies. The features of this module may include:
- Profile Creation: A student can create a profile that includes their personal details, education qualifications, skills, achievements, and work experience.
- Job Search: Students can search and filter job opportunities based on their preferences such as job type, location, salary, and industry.
- Apply for Jobs: Once a student finds a job opportunity that matches their qualifications, they can apply for the job by submitting their application and resume.
- Interview Scheduling: If the student's application is accepted, they may receive an invitation for an interview, which they can accept and schedule through this module.

TPO Module:

- The Training and Placement Officer (TPO) module is designed to manage the placement process for the college or university. The features of this module may include:
- Company Registration: A TPO can register companies that are interested in recruiting students from the institution.
- Job Posting: A TPO can post job opportunities from registered companies and manage them, including setting criteria for students to apply.
- Student Placement Management: The TPO can track the placement of students in various companies and industries, including updating their status and feedback from recruiters.

Admin Module:

- The admin module is designed to manage and control the overall functioning of the placement portal. The features of this module may include:
- User Management: An admin can manage the user accounts, including students, TPOs, and recruiters.
- Content Management: An admin can manage the content on the portal, including job postings, events, and notifications.

- Reporting and Analytics: An admin can view reports and analytics related to job postings, student applications, placements, and more.

Our platform leverages Web Scraping with Python and BeautifulSoup to extract company information, reviews, and

interview questions directly from Glassdoor. Using Linear Regression Natural Language Processing algorithms, we process reviews to assign sentiment scores, resulting in a final cumulative score for each company. Our system also utilizes advanced DBMS concepts such as Query Scheduling, Transactions, and Crons for increased flexibility, speed, and stability. Additionally, the platform is hosted on AWS, ensuring both power and security.

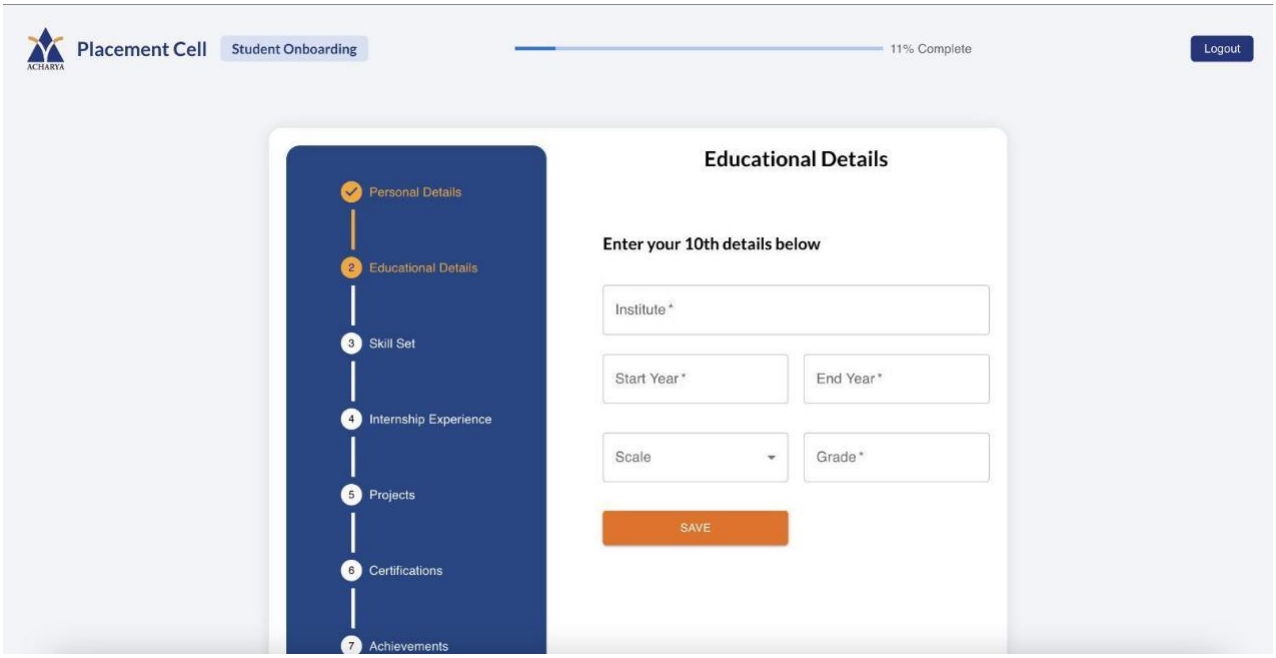

Figure 5.1: Student profile building onboarding

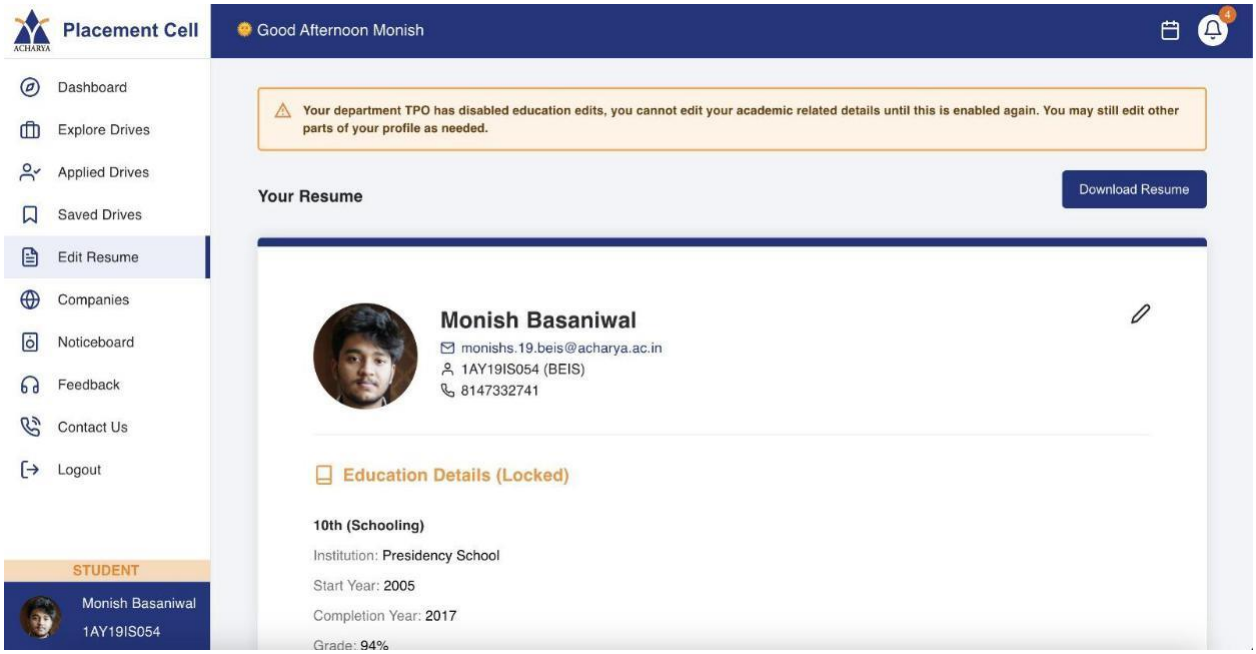

Figure 5.2: Student resume with edit access restricted

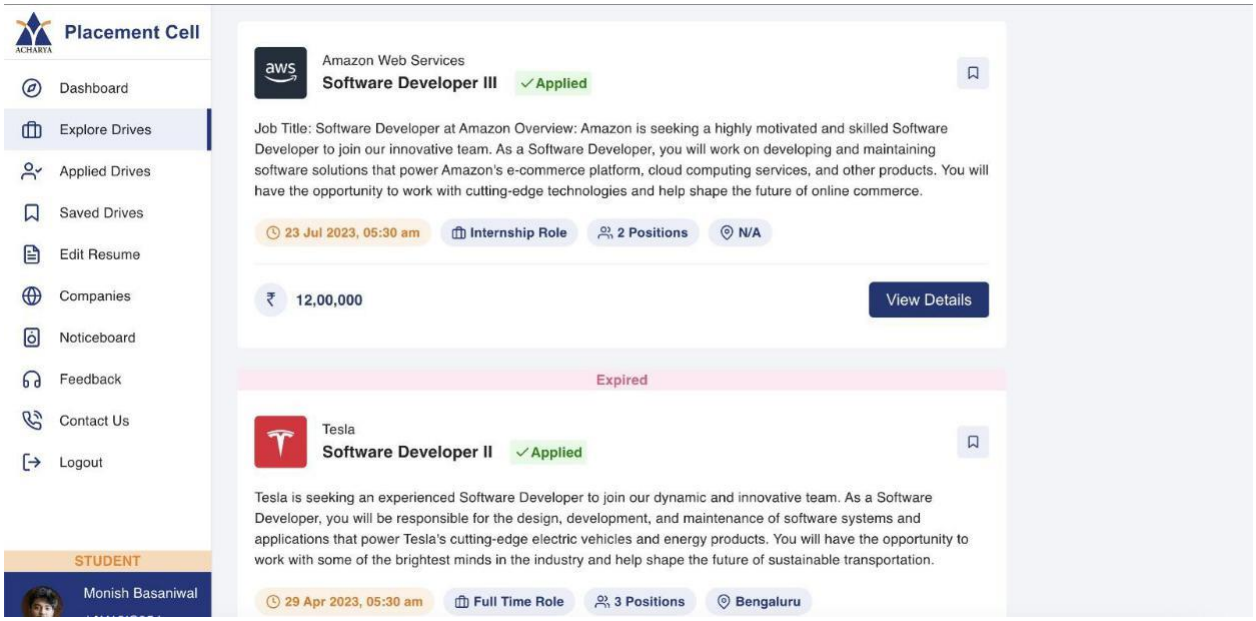

Figure 5.3: Student drives listings

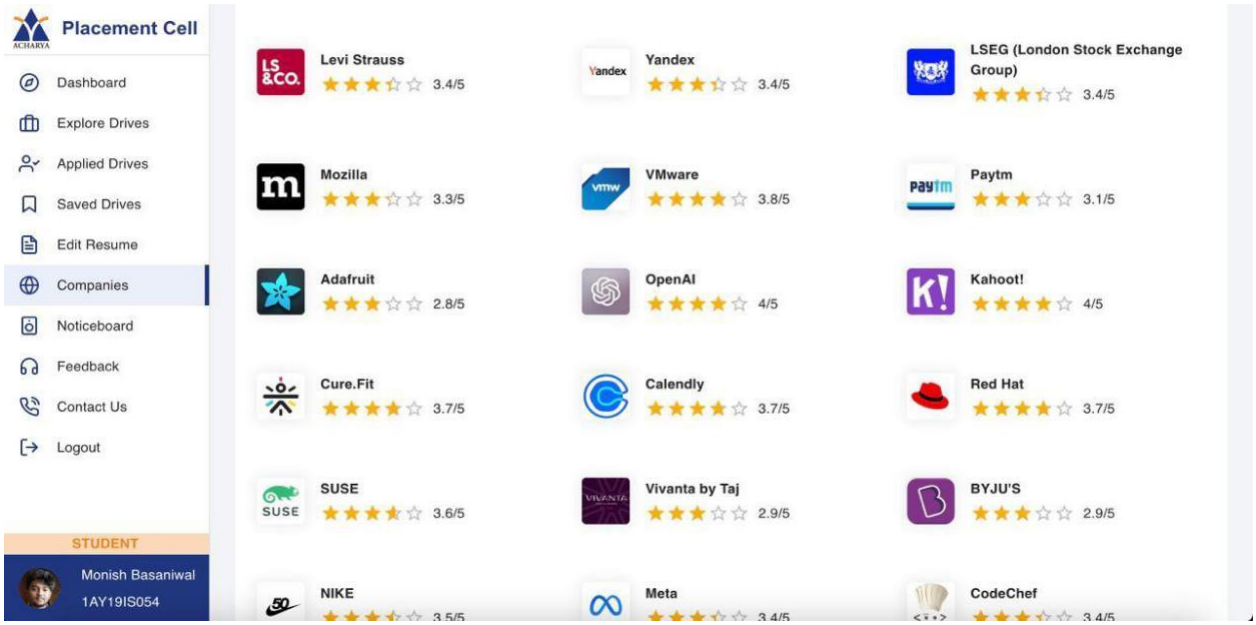

Figure 5.4: Whitelisted companies list with ratings

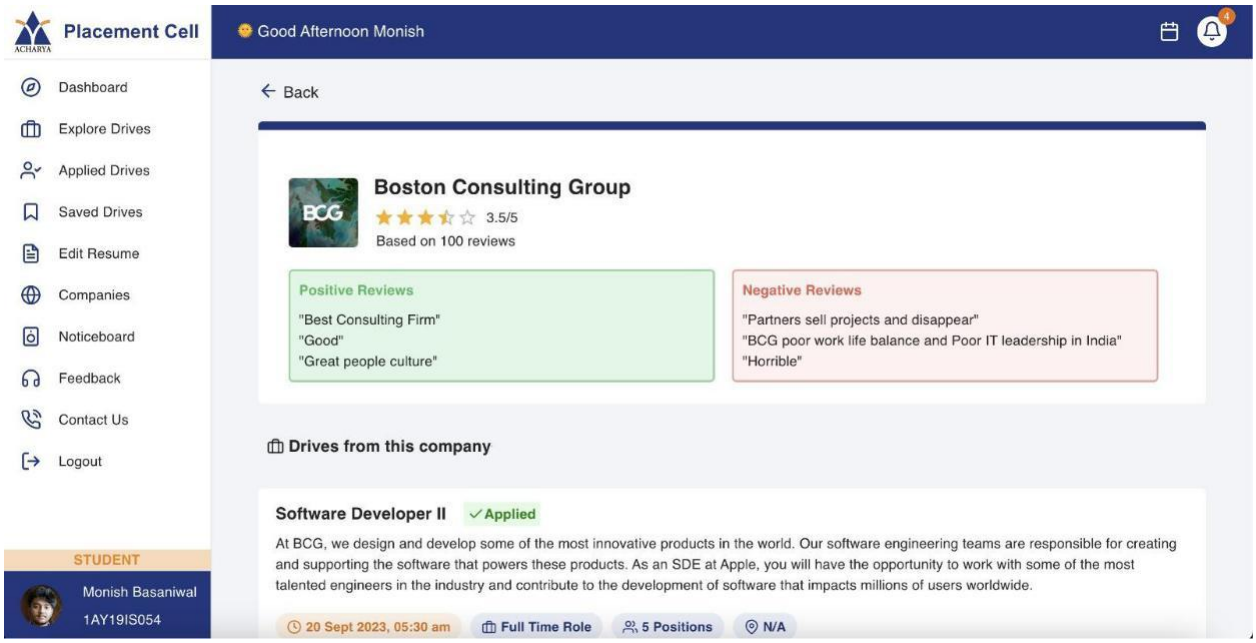

Figure 5.5: Specific company with Pros and Cons mentioned

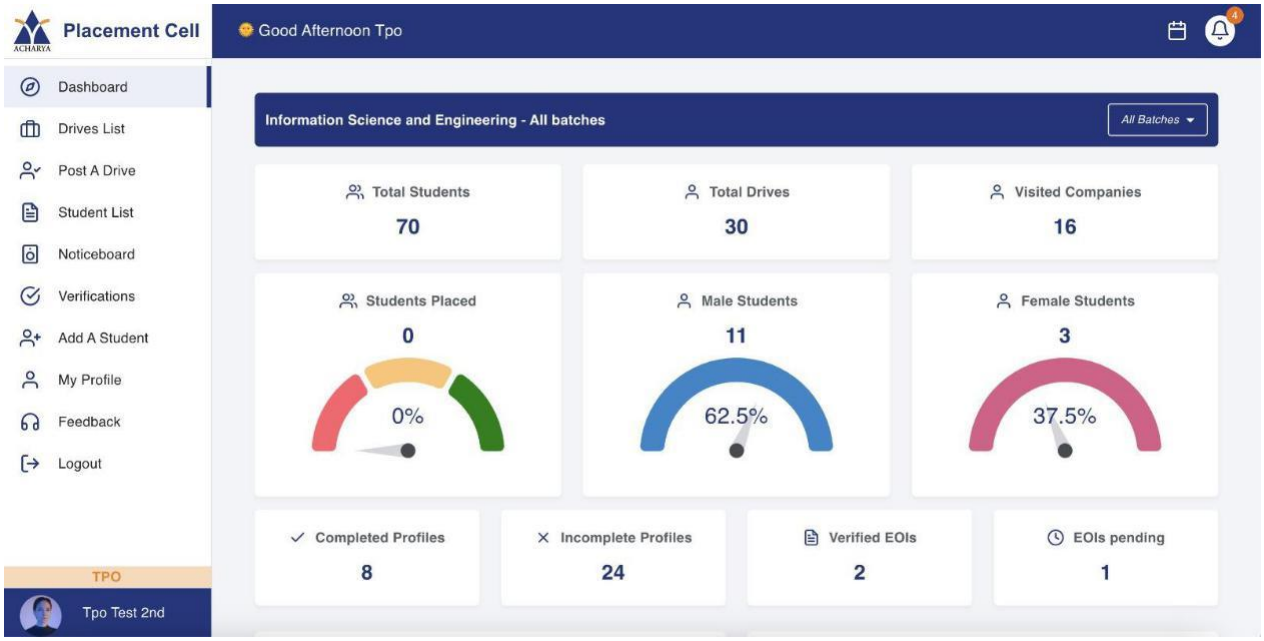

Figure 5.6: TPO Dashboard with reports

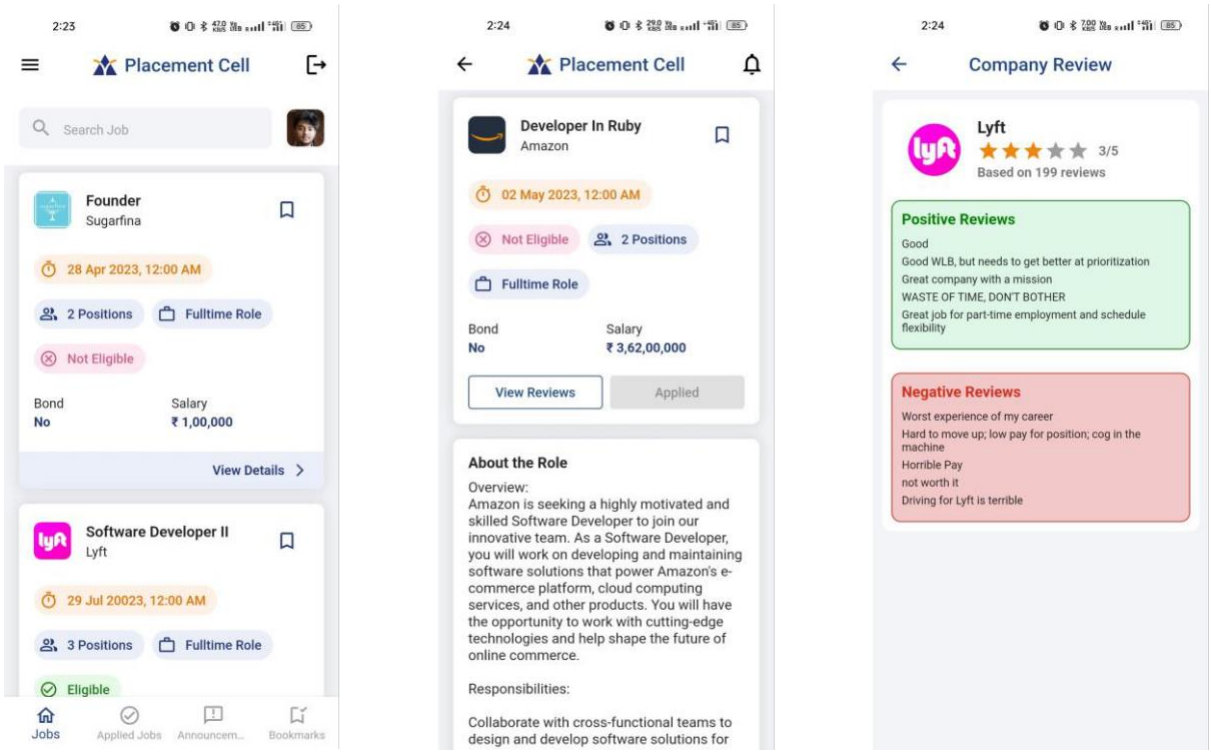

Figure 5.7: Student module on mobile app

Conclusion and Future Enhancement

For students, these modules have significantly enhanced their experience by offering user-friendly interfaces and intuitive features. Students can now easily search and apply for relevant job opportunities through a single platform, saving them time and effort that would have been spent on manual searches. The modules may include features such as personalized job recommendations, resume building tools, and interview preparation resources, which empower students to present themselves effectively to potential employers.

TPOs play a crucial role in facilitating the placement process, and the implementation of dedicated modules has transformed their operations. These modules enable TPOs to efficiently manage and coordinate various activities, including job postings, campus interviews, and placement drives. By automating these processes, TPOs can focus more on building relationships with employers, engaging in career counselling with students, and organizing skill development programs. The modules may also provide data analytics and reporting features that help TPOs track the progress and success rates of placements, enabling them to identify areas for improvement and make data-driven decisions.

From an administrative perspective, the implementation of these modules offers a centralized and comprehensive solution for managing the entire placement process. Administrators can have real-time access to data and insights related to job postings, student applications, and placement outcomes. This enables them to monitor and track the progress of placements, ensuring compliance with placement policies and regulations. Administrators can also generate reports, perform data analysis, and gather valuable feedback from students, TPOs, and employers, which helps in refining and optimizing the placement process over time.

In conclusion, the implementation of student, TPO, and admin modules has revolutionized the placement process for colleges and universities. These modules

have significantly simplified and streamlined job searching and application processes for students, facilitated efficient management of job opportunities and placement processes for TPOs, and provided a comprehensive and centralized platform for admins to oversee and optimize the entire process.

Looking ahead, there is vast potential to deploy the college placement software at

the institute level, further enhancing the management of student placements and training. Future endeavours could involve expanding the software's capabilities to encompass additional features such as skill assessments and personalized career guidance. Moreover, integrating the software with other institutional systems would create a seamless experience for students and staff, bolstering efficiency and effectiveness. With continued development, the college placement software has the potential to become a cornerstone of the institute's career development and placement services, enabling students to achieve their career aspirations and assisting the institution in producing a highly qualified and competent workforce.