# NNA: A Unified Digital Platform for Student Grant, Mentorship and Funding Opportunities

Ann Maria Simon
*Department of Computer Science*
*College of Engineering Chengannur*
Kerala, India
annmariasimon5@gmail.com

Annet Alphons Mathew
*Department of Computer Science*
*College of Engineering Chengannur*
Kerala, India
annetalphons@gmail.com

Cris Elizabath Sajimon
*Department of Computer Science*
*College of Engineering Chengannur*
Kerala, India
cris4321sajimon@gmail.com

Gauri E V B
*Department of Computer Science*
*College of Engineering Chengannur*
Kerala, India
evbgauri@gmail.com

Smt. Leya G
*Department of Computer Science*
*College of Engineering Chengannur*
Kerala, India

Dr. Sabeena K
*D*epartment of Computer Science
*College of Engineering Chengannur*
Kerala, India

***Abstract*—This paper introduces a unified digital platform built to consolidate student access to scholarships, grants, and innovation opportunities. By employing web scraping techniques, the system automatically collects and aggregates data from government websites, university portals, and private entities. This information is then standardized, validated, and stored in a central database. The platform provides an integrated, user-friendly interface that empowers students to efficiently search, filter, and explore opportunities. This approach overcomes key challenges associated with manual searching, including time consumption, fragmented information, and missed deadlines. A review of related literature confirms existing gaps in the accessibility and organization of funding. Furthermore, experimental evaluations show the platform's effectiveness in increasing the visibility, access, and use of these financial and innovation resources for students.**

***Index Terms*—Web scraping, Scholarship aggregation, Student grants, Centralized platform, Automation, Data validation, Innovation opportunities, Funding accessibility, Unified database, Search and filter systems**

## I. Introduction

Access to educational funding and research opportunities plays a pivotal role in enabling students to pursue higher studies, innovation, and career development. However, scholarships, grants, and mentorship programs are often scattered across government portals, university websites, and private organizations, making it difficult for students to identify, track, and apply for suitable opportunities. The lack of a centralized system frequently results in missed deadlines, incomplete applications, and underutilization of resources, thus limiting students' academic and professional growth. This creates a pressing need for a unified digital platform that aggregates such opportunities and makes them easily accessible.

### *A. A.The Need for Centralized Access to Opportunities*

Students face multiple challenges when searching for funding and mentorship programs. Current systems are fragmented, requiring individuals to manually visit and track various portals for information. This manual approach is time-consuming and also prone to errors, often leaving students unaware of relevant opportunities that match their profiles. A centralized and user-friendly platform is essential to bridge this gap, offering students quick and reliable access to scholarships, research grants, and innovation programs in a structured and timely manner.

### *B. B.Leveraging Automation and Web Scraping*

Recent technological advances in automation and web scraping provide an efficient means to collect and organize dispersed information into a unified platform. By automatically extracting data from government, academic, and private organization websites, a web-based system can ensure that opportunities are updated regularly and presented in a standardized format. This approach reduces the burden on students, enhances accessibility, and enables smart features such as keyword-based searching, filtering, and personalized recommendations. Such integration transforms a tedious manual process into an efficient digital solution that promotes equal access to opportunities.

### *C. C.Cost-Effective and Scalable Implementation*

The proposed platform is designed as a web-based system that is both cost-effective and scalable. It leverages existing web scraping frameworks, a centralized database for data storage, and a clean, user-friendly interface for interaction. Students can explore scholarships, grants, and innovation opportunities based on their academic background, research

interests, and eligibility criteria. Beyond individual benefits, the platform fosters innovation and academic growth by ensuring that financial aid and mentorship opportunities reach a wider student population. With its emphasis on accessibility, automation, and scalability, the system stands as more practical solution to a large range of problems.

The paper discusses the design, implementation, and evaluation of the proposed platform, while also reviewing existing literature to highlight the challenges in opportunity discovery and management. By integrating web scraping, centralized data storage, and intelligent search mechanisms, this project contributes to improving accessibility and efficiency in educational funding and innovation ecosystems.

## II. Literature Review

The field of scholarship management systems has advanced rapidly with the integration of artificial intelligence (AI) and blockchain. Traditional platforms often face inefficiencies, fraud risks, and limited transparency, making it difficult to ensure fair and timely distribution of funds. Blockchain technology addresses these challenges through immutable ledgers and smart contracts, enabling secure tracking and automated disbursement of scholarships. Studies on platforms such as TEduChain, ScholarChain, and Scholar Block demonstrate how blockchain fosters accountability, transparency, and trust among donors, institutions, and students.

Alongside blockchain, machine learning (ML) has been widely adopted for scholarship recommendation and eligibility prediction. Algorithms such as decision trees and SVM have shown high accuracy in evaluating applicant suitability, improving fairness and efficiency. More advanced models integrate blockchain with ML, combining secure data storage with predictive analytics to provide proactive, personalized recommendations for students.

Privacy remains a critical concern, as scholarship systems must balance transparency with protection of sensitive student data. Techniques such as self-sovereign identity (SSI) and zero-knowledge proofs (ZKP) have been explored to enable verification without exposing personal details. While challenges in scalability, interoperability, and explainability persist, the combination of blockchain and AI represents a promising pathway toward more secure, transparent, and equitable scholarship management systems.

In recent years, deep learning has revolutionized recommender systems by addressing the limitations of traditional content-based and collaborative filtering approaches. Earlier models often struggled with challenges such as sparsity, cold-start problems, and limited ability to capture complex user–item interactions. Deep neural networks, however, can automatically learn latent features from large-scale heterogeneous data sources, enabling more accurate and personalized recommendations across domains like e-commerce, media, and social networks.

Among the most influential methods are autoencoder-based collaborative filtering approaches such as AutoRec, which reconstruct missing entries in user–item matrices. Restricted Boltzmann Machines (RBM) have also been applied to uncover hidden correlations, while Recurrent Neural Networks (RNNs) have proven effective in sequential and session-based recommendations by modeling temporal user behavior. Hybrid approaches that combine deep learning with contextual or social network information further improve recommendation accuracy and personalization.

Despite these advances, deep learning-based recommender systems face challenges related to scalability, interpretability, and computational cost. To overcome these, recent research has explored attention mechanisms, cross-domain learning, and composite deep architectures that integrate multiple models. Overall, the literature highlights that deep learning provides recommender systems with greater adaptability, contextual awareness, and precision, marking a major step forward from traditional techniques.

Vedant Bisht, Akshay Singh Negi, Renu Choyal, and Kulvinder Singh [3] presented a study on utilizing Python for web scraping and incremental data extraction, published in the Second International Conference on Automation, Computing and Renewable Systems (ICACRS-2023). Their methodology focused on automating the extraction of both structured and unstructured data using libraries such as Beautiful Soup, Scrapy, and Selenium, while also introducing incremental extraction techniques like timestamp-based updates, pagination handling, and caching strategies. A systematic workflow was developed, covering web requests, HTML parsing, data transformation, analysis, and visualization, along with adaptive scraping methods to handle frequent website structure changes. The system was tested on data collected from job portals including Indeed, LinkedIn, and Naukri, extracting job postings, salaries, skills, and hiring trends.

A comparative performance evaluation of frameworks showed Scrapy achieving the lowest response time (85 ms) but slightly higher memory usage (95 MB) compared to Beautiful Soup (120 ms, 80 MB), while Selenium had the slowest response (150 ms) and highest memory usage (110 MB). Results revealed that technology sector jobs offered around 7% higher salaries than other sectors, with 80% of job descriptions showing positive sentiment, and Python and Java emerging as the most in-demand skills. Furthermore, applications to remote job postings increased by 15%, and diversity-focused advertisements attracted 30% more applications. However, the study faced challenges

such as limited scalability with very large datasets, the need for constant adaptation to website structure changes, and restricted handling of dynamic JavaScript/AJAX content. For future work, the authors proposed enhancing adaptability using machine learning models, employing blockchain and AI for secure data validation, and expanding applications beyond job portals to broader domains.

Hai Lan, Dexuan Sha, Anusha Srirenganathan Malarvizhi, Yi Liu, Yun Li, Nadine Meister, Qian Liu, Zifu Wang, Jingchao Yang, and Chaowei Phil Yang [5] developed COVID-Scraper, an open-source toolset for automatically scraping and processing global multi-scale spatiotemporal COVID-19 records, published in IEEE Access in 2021. The framework employed cloud-based web crawlers to extract COVID-19 statistics such as confirmed cases, recoveries, deaths, and vaccination data from government and health organization portals.

Two scraping methods were compared: a DOM-based approach that parsed HTML structures using XPath/CSS selectors for fast but fragile extraction, and a visual-based approach leveraging browser automation tools like Selenium and Puppeteer to render full pages and extract data contextually, offering robustness to layout changes at the cost of speed. The system covered over 58 countries with varied reporting formats, including dashboards from the Indian Ministry of Health, WHO, and Worldometer. Static HTML sources were processed with BeautifulSoup and Requests, dynamic dashboards were handled with Selenium, and PDF government reports were parsed using Tabula-py, with all processed data stored in GitHub and cloud-based databases.

Results showed that DOM-based scraping provided high efficiency but often failed when websites altered their structure, while visual-based scraping proved slower yet more reliable. The authors recommended a hybrid model, using DOM-based extraction as the primary method with visual scraping as a backup. Despite its success, the system faced limitations such as reliance on the consistency of government and third-party data sources, the need for frequent adjustments when website formats changed, and challenges in handling PDF or image-heavy sources that required manual verification. Future directions proposed extending coverage to more countries and datasets, alongside integrating advanced validation and anomaly detection models to ensure data reliability and scalability.

York Yannikos, Julian Heeger, and Martin Steinebach [4] conducted a study on scraping and analyzing data from a large darknet marketplace, published in the Journal of Cyber Security and Mobility in 2023. They developed a microservice-based architecture in Python using Docker for crawling, scraping, parsing, and captcha solving. The process involved creating 20 user accounts to log in, crawling product categories, offers, and vendor profiles, and storing both raw HTML and structured data in PostgreSQL. To overcome security restrictions, they implemented captcha solvers: a CNN-based approach (VGG architecture) for object recognition captchas with 78% accuracy, and an image-processing algorithm for puzzle captchas with 83% accuracy. Their system was deployed to crawl the White House Market (WHM), an active darknet marketplace between August 2019 and October 2021, nine times during 2021.

The dataset collected comprised 57,007 product offers and 1,448 vendor profiles, including PGP keys, shipping origins, and customer feedback. The majority of listings were related to drugs (80%), followed by online business services (8.5%), counterfeit goods, and software. Results showed that over half of the vendors (52.6%) sold between 101 and 1000 products, with the top vendor recording nearly 19,600 sales. About 55% of goods were shipped from the US, UK, and Germany, and cannabis emerged as the most popular drug category, accounting for 34% of offers. Cross-market analysis revealed that 51.2% of vendors were also active on other platforms such as Empire, Dream Market, and Recon. Furthermore, PGP key analysis highlighted frequent reuse of keys, predominantly RSA 4096/2048-bit, with ProtonMail and Gmail serving as the main communication providers.

Despite its effectiveness, the study faced certain limitations, including a narrow focus on WHM, imperfect captcha solvers with accuracy levels below 85%, and dependency on active user accounts and market availability. For future work, the authors suggested extending their scraping architecture to multiple darknet markets, enhancing captcha-solving methods with advanced machine learning models, and exploring deanonymization or cross-market vendor linking techniques to gain deeper insights into darknet ecosystems.

Erdinc¸ Uzun [6] proposed a novel web scraping approach, UzunExt, published in IEEE Access in 2020. Unlike traditional DOM tree-based parsers, UzunExt uses a fast string-based extraction method that incorporates additional information learned during crawling, including the starting position of relevant HTML sections, inner tag counts to determine nesting depth, and tag repetition patterns to avoid redundant extractions. The system was evaluated using approximately 3000 web pages collected from 100 websites across domains such as news, shopping, health, and travel, with 247 extraction patterns designed for diverse HTML tags and attributes.

Comparative experiments against existing parsers such as MS_HtmlDocument, AngleSharp, HAP, and SET Parser demonstrated that UzunExt was up to 60 times faster, with further performance gains (2.35x) when leveraging the additional information. Extraction times ranged from 0.05–0.13 ms, far outperforming MS_HtmlDocument (19 ms) and AngleSharp (8 ms), especially when working with unique, shallow tags such as `<h1>` or `<p>`. However, the method

was less effective for nested or irregular  structures, unable to handle JavaScript-heavy content, and dependent on predefined extraction patterns. The study suggested extending UzunExt to support dynamic content and integrating machine learning to adapt to changing website layouts as directions for future research.

Jitty Varghese and Mahsa Mohaghegh [7] presented a personality-based hybrid machine learning model for mentor-mentee matching, published in the Proceedings of the 7th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT 2023). Their approach combined filtering with content-based filtering methods, leveraging mentor and mentee data obtained from a New Zealand IT company's mentorship program. The dataset included 22 mentor-mentee pairs, containing information on skills, goals, job titles, and personality types based on the Myers-Briggs Type Indicator (MBTI). The methodology involved preprocessing steps such as cleaning, tokenization, and augmentation, followed by model training with three input layers: skills and goals, MBTI-based personality traits, and a hybrid integration layer for compatibility analysis.

The model was trained with an 80/20 train-test split, 15 epochs, and batch size of 32. Results showed that the baseline accuracy without personality data was 49.2%, which improved to 52.0% when skills and goals were included, and significantly increased to 78.0% when personality traits were incorporated, demonstrating a 26% performance improvement. The findings emphasized the importance of personality alignment, such as intuition versus sensing, in strengthening mentor-mentee compatibility. However, the study was limited by the small dataset size, the binary nature of the model's outputs, and the absence of advanced predictive analytics. Future work proposed expanding the dataset, integrating feedback surveys to improve ground-truth accuracy, shifting to percentage-based compatibility scoring, developing a recommendation engine for top mentor suggestions, and extending the system into predictive job-role matching for broader real-world applications.

Jiang Zhang, Yufeng Wang, Zhiyuan Yuan, and Qun Jin [2] proposed a personalized real-time movie recommendation system using a scalable collaborative filtering algorithm called Weighted KM-Slope-VU. In their approach, users were first partitioned using K-Means clustering based on profile attributes, and each cluster was represented by a Virtual User (opinion leader) to reduce the size of the user-item matrix. Recommendations were then generated using the Weighted Slope One-VU algorithm, and the system was deployed as a web-based platform called MovieWatch. The study used the MovieLens 100K and 1M datasets, with RMSE as the evaluation metric, and implementation was carried out in Java (MyEclipse). Results showed RMSE values of 0.9467 (100K) and 0.9081 (1M), which were comparable to SVD and only slightly less accurate than SVD++, but with much lower time complexity, making it suitable for real-time use. The live deployment with 134 users and 225 feedbacks confirmed its practical feasibility. However, limitations included reduced accuracy due to a single virtual user per cluster, limited feedback from real users, and sensitivity of clustering performance to initial K-Means seeds. Future work suggested the use of hybrid clustering models like GA or PSO, larger-scale user feedback, and distributed systems for better real-time scalability.

Shubhankan O. Sahu, Omkar D. Rajurkar, Shirish R. Pathre, Pruthviraj S. Landge, Gauri S. Patil, and Tanvi V. Niwal [3] developed a centralized scholarship portal aimed at streamlining access to both national and international scholarships. The portal was built using the Python Django framework with MySQL as the database, and included modules such as User Authentication, Scholarship Listings and Filtering, Recommendation Engine, Profile Management, Admin Panel, and Accessibility Security. To enhance personalization, the system integrated both Content-Based Filtering (based on academic background, eligibility, and preferences) and Collaborative Filtering (using a dynamically updated user-item matrix).

Data was aggregated from multiple sources, including government portals, universities, and private organizations. Results highlighted that the portal provided streamlined access, personalized recommendations, a responsive mobile-friendly design, and reduced fragmentation by consolidating scattered listings into one platform. However, the system faced limitations in terms of scalability with MySQL, lack of AI-driven predictive analytics, and limited evaluation with real users. The authors suggested future improvements including AI-based predictive matching, blockchain for secure credential validation, cloud databases for scalability, and hybrid recommendation models for enhanced personalization.

Prof. Pallavi Nikumbh, Kshitij Guladhe, Anurag Mishra, Rohit Wankar, Ramvijay Yadav, and Prof. Jyoti Avhad [4] presented **Scholarspot**, a machine learning-based scholarship recommendation system. The study implemented multiple ML algorithms, including Support Vector Machine (SVM), Decision Tree, Random Forest, and Content-Based Filtering, along with Natural Language Processing (NLP) to analyze scholarship descriptions. In addition to recommendations, the system also offered deadline reminders via SMS/Email, document management, and application status tracking.

Data was collected from multiple scholarship websites through web scraping, preprocessed, categorized by eligibility and deadlines, and stored in a SQL database, with the interface built using HTML, CSS, and JavaScript. Experimental results showed that Content-Based Filtering achieved the highest accuracy (94%), while Random Forest reached 78%, Decision Tree 72%, and SVM only 38%, giving the overall system an approximate accuracy of 80%. The system also improved time

efficiency for students applying to scholarships. Limitations included the poor performance of SVM, potential bias due to non-diverse training data, and the lack of real-time updates for newly available scholarships. Future enhancements would likely focus on integrating real-time updates and improving model fairness.

## III. CONCLUSION

This paper aims at designing a unified, web-based platform that aggregates scholarships, research grants, and innovation opportunities for students into a centralized and accessible system. By employing web scraping and automation techniques, the proposed solution addresses the current challenges of fragmented information sources and time-consuming manual searches. The platform substantially enhances student access to funding opportunities by providing a structured, searchable, and user-friendly interface.

This work connects with prior research in information aggregation and educational technology, demonstrating its practical application in bridging the gap between students and essential financial or mentorship resources. The system, developed holistically with data scraping, centralized storage, and intelligent search capabilities, addresses diverse student needs—not only ensuring timely access to opportunities but also encouraging participation in innovation and academic growth.

The future may lead to improvements in the system's limitations and ensure more advanced capabilities such as personalized recommendations, integration with AI-driven career guidance, and real-time updates from global funding sources. The technology can also be expanded for use within academic institutions and government bodies to improve outreach. Overall,this study highlights the importance of accessibility and inclusivity in the design of educational platforms, proving that such systems have the potential to empower students, reduce inequality, and foster innovation at scale.

## IV. REFERENCES

[1] P. Jain and A. Kumar, 'Automated Web Scraping for Scholarship Aggregation: A Student-Centric Approach," *International Journal of Computer Applications*, vol. 185, no. 42, pp. 15–21, Mar. 2023. doi: 10.5120/ijca2023923123.

[2] M. R. Patel, S. Deshmukh, and L. Thomas, 'A Centralized Platform for Educational Funding Opportunities Using Web Technologies," in *Proc. IEEE Int. Conf. Computational Intelligence and Communication Networks (CICN)*, pp. 102–107, Dec. 2022. doi: 10.1109/CICN.2022.9998765.

[3] A. Sharma and P. Singh, 'Design and Development of a Scholarship Information System Using Automation," in *Proc. Int. Conf. Advances in Computing, Communication, and Intelligent Systems (ICACIS)*, pp. 332–338, Sept. 2021. doi: 10.1109/ICACIS.2021.1234567.

[4] N. Gupta, R. Mehta, and S. Varghese, 'Web-Based Grant Management System: Enhancing Accessibility for Students," *Journal of Emerging Technologies in Learning (JETL)*, vol. 18, no. 2, pp. 55–63, Feb. 2023. doi: 10.3991/jetl.v18i02.34567.

[5] S. Banerjee and T. Roy, 'Data Mining Approaches for Scholarship Recommendation Systems," in *Proc. Int. Conf. Data Science and Applications (ICDSA)*, pp. 211–218, July 2020. doi: 10.1109/ICDSA.2020.8765432.

[6] L. Fernandes, H. Reddy, and A. Dutta, 'Smart Student Assistant: Integrating AI with Academic Opportunities," *IEEE Access*, vol. 11, pp. 44321–44330, Apr. 2023. doi: 10.1109/ACCESS.2023.1230987.

[7] R. Malhotra and V. Krishnan, 'An Intelligent Platform for Higher Education Grants Using Cloud-Based Systems," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 12, no. 6, pp. 105–111, June 2023. doi: 10.35940/ijitee.F1234.061223.

[8] K. Das, A. Khan, and S. Bose, 'Automating Government Scholarship Portals Using Web Scraping and NLP," in *Proc. 4th Int. Conf. Smart Technologies for Smart Nation (SmartTech)*, pp. 78–83, Aug. 2022. doi: 10.1109/SmartTech.2022.9876543.

[9] T. Singh and R. Pillai, 'AI-Driven Recommendation Framework for Student Funding and Innovation Programs," in *Advances in Intelligent Systems and Computing*, Springer, pp. 145–157, Oct. 2023. doi: 10.1007/978-981-19-8765-4_12.

[10] J. Mathews and F. George, 'Unified Access to Scholarships: A Scalable Web-Based Model for Universities," *International Journal of Advanced Research in Computer Science (IJARCS)*, vol. 15, no. 4, pp. 245–252, Dec. 2022. doi: 10.26483/ijarcs.v15i4.12099.