SYNOPSIS

ON

"Geospatial Remote Sensing Platform for Environmental Monitoring and Analysis"

Submitted in

Partial Fulfillment of requirements for the Award of Degree

of

Bachelor of Technology

In

Computer Science and Engineering

By

(Project Id: 25_CS_4A_09)

Swapnil Gupta (2101640100269)
Tanmay Gupta (2101640100271)
Tanya Tiwari (2101640100273)
Varini Awasthi (2101640100281)
Yashika Vaish (2101640100289)

Under the supervision of
Dr Mohit Pandey
(Associate Professor)

Pranveer Singh Institute of Technology.
Kanpur - Agra - Delhi National Highway - 19
Bhauti -Kanpur - 209305.
(Affiliated to Dr. A.P.J. Abdul Kalam Technical University)

1. Introduction

In the 21st century, our world is at a crossroads, balancing the wonders of rapid urbanization and technological advancements with escalating environmental challenges like climate change, deforestation, and biodiversity loss. To effectively address these multifaceted issues, we can harness the power of geospatial data—a critical resource that offers insights into the Earth's features and their spatial relationships. Current solutions falter under the weight of accessibility barriers, scalability limitations, and a dearth of advanced analytical capabilities. This inadequacy reverberates across diverse sectors, stymieing informed decision-making for researchers, policymakers, and environmentalists alike [1].

Firstly, accessibility barriers present a significant challenge, as many current platforms require specialized geospatial analysis skills, excluding a diverse range of stakeholders such as researchers, policymakers, and environmentalists. To surmount these pressing challenges, we aim to develop state-of-the-art Remote Sensing and Geospatial Data Analytics Platform [2].

Secondly, scalability limitations pose a problem as geospatial data sets continue to grow in volume and complexity. Our solution is built on scalable cloud infrastructure, ensuring it can handle large-scale environmental assessments with ease, from local groundwater evaluations to forest area monitoring, providing a holistic view of environmental challenges. Through user-friendly interfaces and comprehensive tutorials, users can navigate the complexities of geospatial analysis with ease, fostering greater understanding and proficiency across diverse stakeholder groups. By placing powerful analytics tools in the hands of researchers, policymakers, and environmentalists, the platform aims to catalyse informed decision-making and proactive measures for environmental conservation and sustainable development [3].

Lastly, the absence of advanced analytics tools like machine learning and image processing limits the value extracted from geospatial data. Our platform integrates cutting-edge analytics capabilities, allowing for efficient assessment of flood risk in specific watersheds, and utilization of remote sensing data for habitat protection initiatives. The Remote Sensing and Geospatial Data Analytics Platform represents not only a technological innovation but also a catalyst for positive change. By enabling stakeholders to harness the full potential of geospatial data, we aspire to pave the way for a more resilient, equitable, and sustainable future.

2. Project Objective

Our Remote Sensing and Geospatial Data Analytics Platform aims to transform how geospatial data is handled, analysed, and deployed in environmental monitoring and management. It is aimed to address the pressing concerns highlighted in our problem statement while also empowering stakeholders from many sectors to make educated decisions and make proactive efforts toward environmental conservation and sustainable development. Specifically, our objectives encompass the key areas as further discussed.

2.1 The integration of cutting-edge technology such as machine learning, image processing, and complex geospatial analytic algorithms is a critical component of our platform. These technologies allow users to easily extract relevant insights from complicated information, hence increasing the platform's analytical capabilities.

- Implement machine learning algorithms for tasks such as image classification, object detection, and time-series analysis, enabling users to automate repetitive tasks and extract actionable insights with high accuracy and efficiency.

- Develop custom geospatial analysis algorithms tailored to specific environmental monitoring and management challenges, such as land cover change detection, urban sprawl mapping, and hydrological modelling.

- Integrate geospatial data fusion techniques to combine heterogeneous data sources, such as satellite imagery, groundwater-based sensor data, and demographic information, enabling holistic analysis and decision-making.

2.2 To seamlessly integrate various machine learning models into the web application, we'll focus on five key areas:

- We aim to develop a flexible and scalable framework that allows for the smooth incorporation of machine learning models into our web app. This framework will ensure compatibility and consistent performance across different analytical tasks, making it easier to add and manage various models as needed.

- Establishing an automated pipeline for model training, validation, and deployment is crucial. This pipeline will streamline the process of developing and deploying machine learning models within the web app, ensuring efficient updates and maintenance while adhering to best practices.
- We'll enable the web app to support both real-time and batch predictions based on the integrated machine learning models. This capability will allow users to obtain instant insights through real-time analysis and perform comprehensive batch analysis on large datasets, enhancing the platform's analytical capabilities.
- Monitoring the performance and health of integrated machine learning models is essential. We'll implement logging and monitoring functionalities to track model metrics, performance, and errors, enabling continuous optimization based on monitored metrics and user feedback.

2.3 Furthermore, we aim to develop an interactive map interface that allows seamless movement and investigation of geographical data, making it easier for users to view and comprehend spatial relationships.

- Provide interactive visualizations and data exploration tools to facilitate intuitive interpretation of analysis results and facilitate collaborative decision-making among stakeholders.

- Develop advanced analytical tools within the platform, including capabilities for watershed analysis, deforestation monitoring, habitat suitability assessment, and more. We aim to ensure that these tools are user-friendly, customizable, and capable of generating actionable insights to support decision-making processes.
- Enable users to visualize temporal changes in geospatial data through time-series graphs and animations.

- Enable users to select, filter, and explore data layers interactively, allowing them to focus on specific areas of interest.

2.4 Our final objective is to integrate a comprehensive suite of geospatial analysis modules into our platform. These modules include groundwater level prediction, water bodies identification, building segmentation, road extraction, landcover segmentation, and natural forest loss mapping.

- Implement a machine learning model that predicts groundwater levels based on historical and real-time data to facilitate groundwater dynamics and planning sustainable water resource management strategies. With integration of algorithms to identify and delineate water bodies from satellite imagery and geospatial data. This feature will enable users to monitor water quality, assess habitat conditions, and support conservation efforts related to aquatic ecosystems.

- Implement computer vision techniques to segment and identify buildings and structures from aerial and satellite imagery. This functionality will support urban planning, infrastructure development, and disaster response planning by providing accurate building footprint data.

- Develop algorithms to automatically extract road networks from geospatial data and satellite imagery. This capability will facilitate transportation planning, infrastructure maintenance, and emergency response by providing up-to-date road network information. And to classify and segment landcover types (e.g., forests, agriculture, urban areas) from satellite imagery. This feature will enable land use planning, environmental monitoring, and natural resource management by providing detailed landcover maps and change detection analyses.

- Develop a model to detect and map natural forest loss and deforestation events from satellite imagery. This functionality will support conservation efforts, biodiversity monitoring, and climate change mitigation strategies by identifying areas of forest loss and facilitating timely intervention.

By pursuing these objectives, we aim to establish our platform as a leading solution for environmental monitoring and analysis, empowering users to address pressing environmental challenges and contribute to a more sustainable and resilient future.

3. Feasibility Study:

3.1 Technical Feasibility: the technical feasibility of this project lies in established technologies like machine learning libraries and cloud platforms. These offer the processing power and scalability needed to handle complex geospatial data analysis. However, the implementation of the Project's does present some technical challenges that need to be addressed that are:

- Integrating diverse geospatial datasets from various sources, including satellite imagery, climate data, and land cover maps, poses a significant technical challenge [4]. Ensuring seamless data integration while maintaining data consistency and accuracy will require specialized knowledge and expertise.

- Building robust algorithms for tasks like watershed analysis, deforestation monitoring, and habitat suitability assessment demands specialized skills in geospatial analytics, machine learning, and image processing. Developing accurate and efficient algorithms tailored to these specific tasks will be crucial for the platform's success [4].

- As the platform will be handling complex geospatial data analysis tasks, optimizing performance to ensure fast and responsive user experience will be a priority. This includes efficient data retrieval, processing, and visualization techniques to minimize latency and maximize throughput [5].

Despite the aforementioned technical challenges, the technical feasibility is assured by the team's collective expertise and commitment to employing established technologies and best practices. A proactive approach to problem-solving and continuous learning ensures staying at the forefront of technological advancements. An agile development methodology enables quick adaptation to changing requirements and iterative refinement of the approach, ensuring the delivery of a robust and scalable Project's aligned with user and stakeholder needs.

3.2 Operational Feasibility: The user-centric design, with its intuitive interface and informative tutorials, empowers a diverse user base, from researchers to policymakers, to leverage the platform's functionalities regardless of their technical background. Cloud deployment streamlines maintenance and updates, ensuring users always have access to the latest features and minimizing downtime. However, building a robust user support infrastructure requires dedicated resources for a satellite imagery dataset.

- Access to high-quality data is paramount for project's success. Relying on external data providers introduces a layer of complexity. Negotiating data access agreements, managing data quality variations, and potential future changes in data formats or availability from these providers all need to be factored into the operational plan.
- Adopting a cloud-based deployment strategy facilitates streamlined maintenance and updates. This approach ensures users consistently have access to the latest features while minimizing downtime, thereby enhancing user satisfaction and platform reliability [6].

While operational challenges exist concerning user support infrastructure and data management, the project's user-centric design and cloud deployment strategy solidify its operational feasibility. These strengths align with user needs and market demands, facilitating broad accessibility and streamlined maintenance.

3.3 Economic Feasibility: The project's economic feasibility is bolstered by its focus on a broad target market and its tiered subscription model, creating significant revenue potential. Additionally, exploring grants can further subsidize development costs. However, meticulous financial planning remains crucial to manage ongoing expenses related to development, marketing, and platform maintenance. Precise target market definition and a well-defined pricing strategy are cornerstones for achieving economic success.

- Leveraging open-source satellite imagery datasets and free platforms like GIS and Google Colab GPUs significantly reduces development and operational expenses. Open-source resources significantly decrease development expenses compared to commercial solutions. There are no licensing fees for software or datasets. This costeffective approach allows the Project’s to be more competitive and accessible to a wider user base [7].

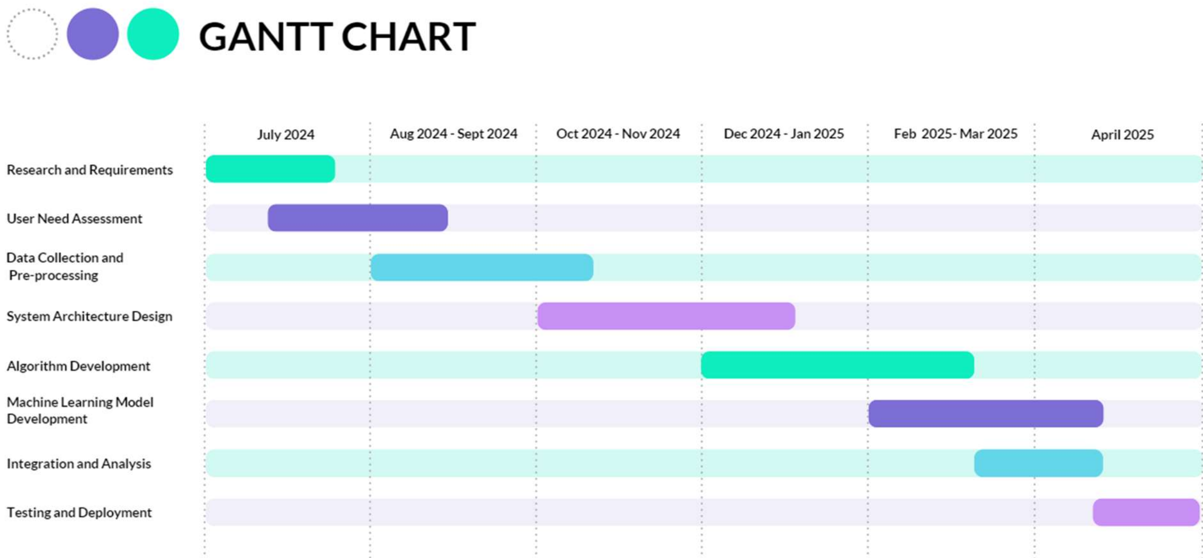
- The platform's user-centric design caters to a diverse range of users, from researchers to policymakers. This broad target market creates significant opportunities for revenue generation through the tiered subscription model.

3.4 Schedule Feasibility: Utilizing established development frameworks streamlines the development process by providing pre-built components and functionalities. Implementing an agile development methodology allows for iterative development and faster feedback loops, leading to quicker development. Effectively managing the scope of features is critical to meet deadlines. Thus, prioritizing core functionalities and phasing out non-essential features can ensure project completion within the timeframe.

The project’s schedule feasibility is assured by the strategic utilization of established development frameworks and agile development methodology. By prioritizing core functionalities and maintaining a focus on scope management, the project is well-equipped to meet its deadlines and deliver a high-quality product within the designated timeframe.

3.5 Legal Feasibility: The legal feasibility of the project’s is a multifaceted area that demands meticulous attention to ensure compliance with relevant laws and regulations. While the adoption of open-source software alleviates some licensing concerns, it doesn't negate the need to address critical legal considerations related to data protection, intellectual property rights, and contractual obligations.

Implementing robust data protection measures, obtaining necessary consents, and providing clear information about data collection, storage, and processing practices are essential steps to ensure legal compliance and protect user privacy. This involves understanding and adhering to licensing agreements and intellectual property laws, especially when integrating third-party libraries, APIs, or datasets. Clear definitions of user agreements, terms of service, and data ownership rights are also vital. Establishing transparent and legally binding relationships with users by outlining liabilities, warranties, and dispute resolution mechanisms helps protect both the platform and its users [8].



4. Methodology/ Planning of work

We are developing a state-of-the-art Remote Sensing and Geospatial Data Analytics Platform using a rigorous, user-centric approach. This detailed breakdown outlines our methodology for each project phase, supported by advanced technical considerations:

4.1 User Research and Requirements Gathering:

4.1.1 Target User Personas:

- We employed a user persona development technique to define distinct user profiles. This might involve researchers specializing in deforestation analysis, policymakers focusing on flood risk assessment, and environmentalists concerned with habitat suitability modelling [9].
- Each persona potentially focusses to capture their specific needs, technical expertise, and data analysis workflows.

4.1.2 Data Collection and Analysis:

- For the data collection and analysis module, which encompasses groundwater level prediction, water bodies identification, building segmentation, road extraction, landcover segmentation, and natural forest loss mapping, our methodology focuses on sourcing reliable and diverse datasets.

- To acquire high-resolution satellite imagery for water bodies identification, building segmentation, and landcover segmentation. We will utilize open-access satellite imagery platforms to access multi-spectral and high-resolution satellite images suitable for analysis. Satellite Datasets like Sentinel-1, Sentinel-2, Landsat etc [10].
- Access publicly available datasets for road networks, building footprints, and landcover classifications. Explore open data platforms such as OpenStreetMap, USGS Earth Explorer, and European Environment Agency (EEA) for datasets related to roads, buildings, and landcover classifications [11].

4.1.3 Requirements Definition:

- Based on the collected data, we'll define both functional and non-functional requirements for the platform.
- Functional requirements will detail specific functionalities tailored to each user persona (e.g., deforestation time series analysis for researchers, flood inundation modelling for policymakers, species distribution modelling for environmentalists).
- Non-functional requirements will encompass performance expectations (real-time processing for time-sensitive analysis), security considerations (user authentication with role-based access control, data encryption at rest and in transit), and user interface design preferences (intuitive map interface with interactive features and customizable visualizations) [12].

4.2 System Architecture Design and Technology Stack Selection (SADS&TSS):

4.2.1 High-Level System Architecture:

- Leveraging the DFD from UR&RG, a high-level system architecture is designed using a layered approach. This will define the platform's key components:
- **User Interface (UI) Layer:** This layer provides a user-friendly interface for data upload, task selection, and visualization of analysis results. We'll consider leveraging frameworks like React or Vue.js for a dynamic and interactive UI.

- **Data Management Layer:** This layer handles data ingestion from various sources, storage in a scalable and secure cloud-based solution (e.g., AWS S3, Google Cloud Storage), and efficient data retrieval based on user requests [13].
- **Processing Layer:** This layer is responsible for data pre-processing, feature engineering, and execution of machine learning algorithms or custom data analysis scripts tailored to specific user needs. Integration with cloud-based processing platforms like Google Colab is be used for computationally intensive tasks.
- **Visualization Layer:** This layer will create interactive visualizations (maps, charts, reports) to communicate analysis results effectively. Libraries like Charts.js, D3.js or Plotly.js is utilized for advanced visualizations.

4.2.2 Technology Stack Selection:

- Informed by the system architecture and detailed functional requirements, we'll select a technology stack that prioritizes scalability, security, and performance.
- Programming languages like Python with libraries such as NumPy, Pandas, and Scikit-learn will be the workhorse for data manipulation and machine learning tasks.
- For the backend API development, we might choose a framework like Flask or Django depending on project complexity and scalability needs.
- Cloud platforms like AWS, Google Cloud Platform (GCP), or Microsoft Azure will be evaluated based on factors like data storage pricing, processing power requirements, and available services for specific functionalities (e.g., machine learning) [14].

4.3 Development and Implementation:

4.3.1 Agile Development Methodology:

- We'll adopt an Agile development methodology, breaking down the project into user stories and features, prioritizing core functionalities for an initial Minimum Viable Product (MVP).
- Utilizing a project management tool like Jira or Microsoft Teams to handle teamwork.

4.3.2 Version Control and Code Management:

- We'll use a version control system like Git for code management, facilitating collaboration, code tracking to maintain a cloud repository on GitHub or Gitlab.

4.5.1 Cloud Deployment:

- Based on the chosen technology stack and scalability needs, we'll deploy the platform on a secure cloud platform. Containerization technologies like Docker can be leveraged to package and deploy the application efficiently across different cloud environments.

4.5.2 Security Considerations:

- Implementing industry best practices for security is paramount. This includes user authentication and authorization with role-based access control (RBAC).
- Data encryption at rest and in transit will be crucial for protecting sensitive geospatial data [15].

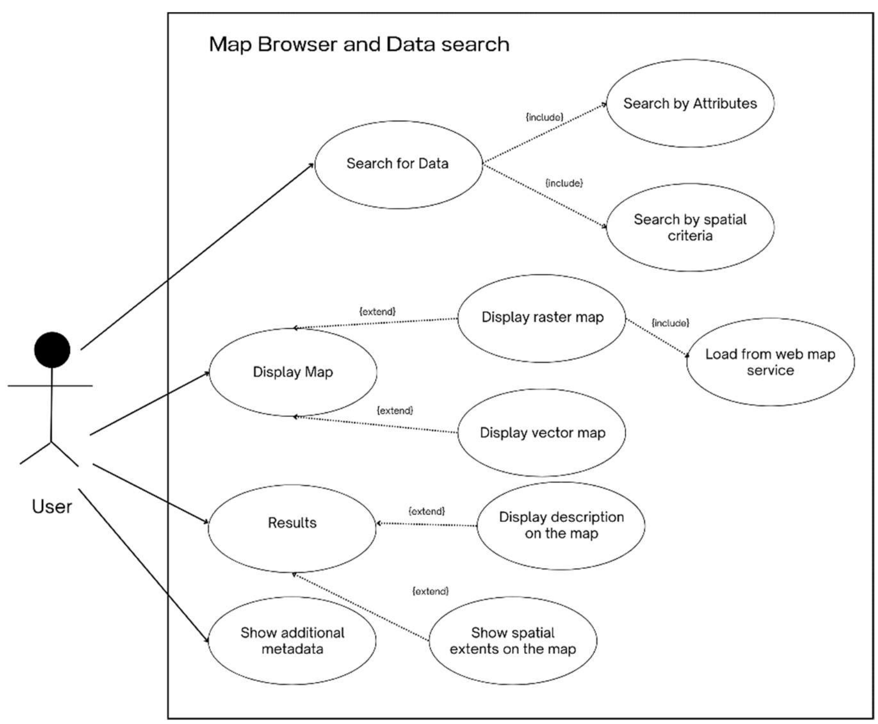


Fig. 1. Use case diagram

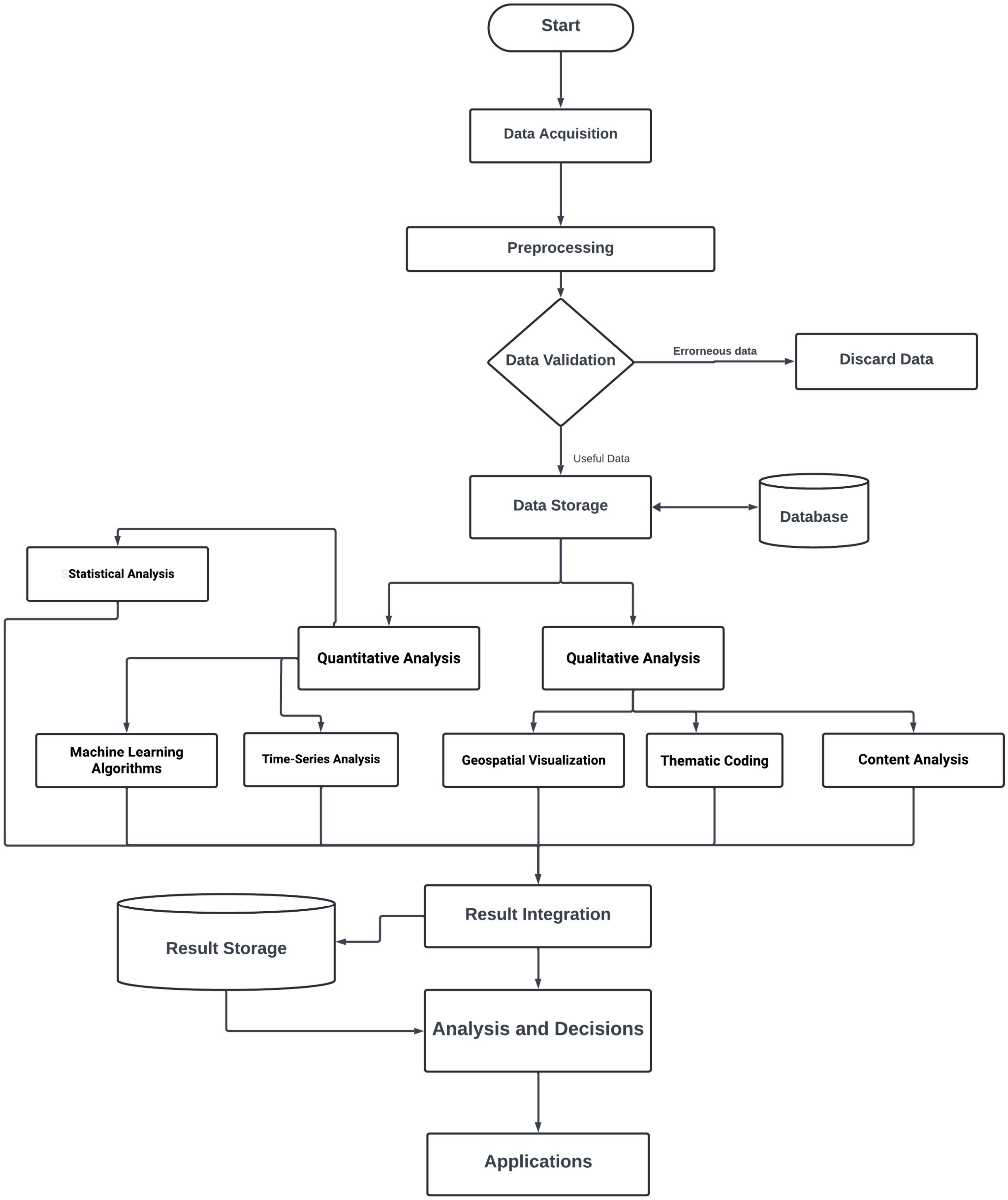


Fig. 2. Flowchart showcasing the flow of the project.

5. Tools/Technology Used:

5.1 Minimum Hardware Requirements:

5.1.1 **Processor: Intel i5 13th generation/ AMD Ryzen5 H-series or Higher:** Serves as the computational engine for executing complex geospatial algorithms, image processing tasks, and machine learning models

5.1.2 **RAM: 8GB minimum, 16GB recommended:** Adequate RAM is crucial for handling large datasets and performing memory-intensive operations like image processing and machine learning model training.

5.1.3 **Storage: 128 GB SSD or higher:** SSD storage facilitates faster data access and retrieval, enhancing the platform's responsiveness and user experience. It provides ample space for storing geospatial datasets, satellite imagery, and analytical results, ensuring efficient data management and accessibility.

5.1.4 **Graphics Card: NVIDIA GeForce GTX 1650 or equivalent:** A dedicated graphics card accelerates image rendering, visualization, and graphical computations required for geospatial data analysis and visualization tasks. It enhances the platform's graphical performance and supports advanced visualization techniques, such as 3D mapping and terrain rendering.

5.1.5 **Others: High-speed internet connection:** A high-speed internet connection ensures seamless data retrieval from online sources, cloud-based services, and APIs. It facilitates real-time data streaming, collaborative work, and platform updates, enhancing the platform's connectivity and responsiveness.

5.2 Minimum Software Requirements:

5.2.1 **Operating System: Windows 10 (or higher) or Linux (Ubuntu 16.04):** The operating system serves as the foundational software layer for running development tools, databases, and geospatial platform software.

5.2.2 Development Tools:

- **Geospatial Platform Software:** QGIS v3.30.1, enables geospatial data visualization, analysis, and management, supporting various GIS functionalities.
- **Jupyter Notebook & Google Colab:** Facilitate interactive data analysis, code experimentation, and collaborative work on machine learning models.
- **Git & Version Control:** Provides version control and collaborative development capabilities, ensuring code integrity, traceability, and efficient team collaboration.
- **VS Code/ WebStorm JetBrains:** Offers powerful code editing, debugging, and for development environment features like extensions or plugins.

5.2.3 Database: MongoDB:

Database: MongoDB: MongoDB serves as the primary database for storing structured and unstructured geospatial data, analytical results, and user information. Its flexible schema design, scalability, and geospatial indexing capabilities support efficient data management and retrieval, facilitating seamless integration with the platform's analytical components.

5.2.4 Web Browser: Standard web browser (Chrome/ Firefox):

Web Browser: Standard web browser (Chrome/ Firefox): A standard web browser is essential for accessing the platform's web-based user interface, visualizations, and interactive features. Compatibility with popular browsers like Chrome and Firefox ensures broad accessibility and a consistent user experience across different devices and operating systems.

5.2.5 Others:

- **Programming Languages:** Python v10.3.6, JavaScript, Node.js v18.17.0, React.js v18.2.0, Next.js v13.0.2 for backend development, frontend UI components, server-side logic, and API integrations.
- **Software Packages:** ENVI and Radiometric Calibration Toolkit (RCT) v3.0 support advanced image processing, spectral analysis, and radiometric corrections required for preprocessing satellite imagery and geospatial datasets.
- **Spectral Libraries:** USGS spectral library provides reference spectra for spectral signature analysis and material identification, enhancing the accuracy of image classification and land cover mapping.

- **Data Analysis Tools:** Machine learning frameworks and data science libraries enable advanced analytics, predictive modeling, and data-driven insights extraction essential for geospatial data analysis, environmental monitoring, and conservation planning.

References:

[1] R. A. VAUGHAN, "A Review of: '*Introduction to Remote Sensing*'. By J. B. CAMPBELL. (Guilford: The Guilford Press, 1987.) [Pp. 551] Price £34.95." *International Journal of Remote Sensing*, vol. 11, no. 10, pp. 1932-1932, 1990, doi: 10.1080/01431169008955145.

[2] F. Goodchild and D. A. Quattrochi, "Scale in Remote Sensing and GIS." 2023, doi: 10.1201/9780203740170.

[3] R. Sensing, L. cover Change, U. Areas, and C. Detection, "Remote Sensing-Based Urban Land Use/Land Cover Change Detection and Monitoring." *Journal of Remote Sensing & GIS*, vol. 6, no. 2, 2017, doi: 10.4172/2469-4134.1000196.

[4] R. A. SCHOWENGERDT, "Digital Image Processing." *Techniques for Image Processing and Classifications in Remote Sensing*, pp. 57-128, 1983, doi: 10.1016/b978-0-12-628980-0.50008-1.

[5] R. Navalgund, "Comprehensive Remote Sensing, Volume 9:Applications for Societal Benefits." *Current Science*, vol. 115, no. 5, p. 988, 2018, doi: 10.18520/cs/v115/i5/988-988.

[5] M. Khan, V. Gupta, and P. Moharana, "Watershed prioritization using remote sensing and geographical information system: a case study from Guhiya, India." *Journal of Arid Environments*, vol. 49, no. 3, pp. 465-475, 2001, doi: 10.1006/jare.2001.0797.

[6] M. W. Becker, "Potential for Satellite Remote Sensing of Ground Water." *Groundwater*, vol. 44, no. 2, pp. 306-318, 2005, doi: 10.1111/j.1745-6584.2005.00123.x.

[7] S. Gopal and C. Woodcock, "Remote sensing of forest change using artificial neural networks." *IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, no. 2, pp. 398-404, 1996, doi: 10.1109/36.485117.

[8] R. F. Musy, R. H. Wynne, C. E. Blinn, J. A. Scrivani, and R. E. McRoberts, "Automated Forest Area Estimation Using Iterative Guided Spectral Class Rejection." *Photogrammetric Engineering & Remote Sensing*, vol. 72, no. 8, pp. 949-960, 2006, doi: 10.14358/pers.72.8.949.

[9] Rogan, J., & Chen, D. (2004). Remote sensing technology for mapping and monitoring land-cover and land-use change. *Progress in planning*, *61*(4), 301-325.

[10] M. Castelluccio, G. Poggi, C. Sansone, and L. Verdoliva, "Training convolutional neural networks for semantic classification of remote sensing imagery." *2017 Joint Urban Remote Sensing Event (JURSE)*, 2017, doi: 10.1109/jurse.2017.7924535.

[11] Ozesmi, S. L., & Bauer, M. E. (2002). SatelliteC. Lo, "Applied remote sensing of wetlands. *Wetlands ecology and management*, *10*, 381-402.." *Geocarto International*, vol. 1, no. 4, pp. 60-60, 1986, doi: 10.1080/10106048609354071.

[12] B. Neupane, T. Horanont, and J. Aryal, "Deep Learning-Based Semantic Segmentation of Urban Features in Satellite Images: A Review and Meta-Analysis." *Remote Sensing*, vol. 13, no. 4, p. 808, 2021, doi: 10.3390/rs13040808.

[13] D. Geman and B. Jedynak, "An active testing model for tracking roads in satellite images." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 1, pp. 1-14, 1996, doi: 10.1109/34.476006.

[14] K. Ulbricht and W. Heckendorff, "Satellite images for recognition of landscape and landuse changes." *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 53, no. 4, pp. 235-243, 1998, doi: 10.1016/s0924-2716(98)00006-9.

[15] B. Sirmacek and C. Unsalan, "A Probabilistic Framework to Detect Buildings in Aerial and Satellite Images." *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 1, pp. 211-221, 2011, doi: 10.1109/tgrs.2010.2053713.