RESEARCH CENTER FOR DATA SCIENCE AND MACHINE LEARNING

-SOFTWARE REQUIREMENT SPECIFICATIONS

SUBMITTED BY:

R SRI VENKAT SAI   211801370080

"RESEARCH CENTER FOR DATA SCIENCE AND MACHINE LEARNING "

A Project report submitted

In the partial fulfillment the award of degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING (2022-2023)

BY

R SRI VENKAT SAI   Reg.No.211801370080

Under the esteemed Guidance of

Mrs. G. Rama Devi, M.Tech,(Ph.d),Asst. Professor

CENTURION UNIVERSITY SCHOOL OF ENGINEERING AND TECHNOLOGY

Rollavaka village, Tekkali mandal 535003

(2022-2023)

CENTURION UNIVERSITY OF TECHNOLOGY AND MANAGEMENT

ANDHRA PRADESH

(2021-2025)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

       This is to certify that the project work entitled **"RESEARCH CENTER FOR DATA SCIENCE AND MACHINE LEARNING"** is a fulfillment of project work done by **R SRI VENKAT SAI (Reg.No.211801370080)** for the award the degree of **BACHELOR OF TECHNOLOGY** in **COMPUTER SCIENCE AND ENGINEERING, CENTURION UNIVERSITY OF TECHNOLOGY AND MANAGEMENT**, during the academic year **2022-2023**.

INTERNALGUIDE   **HEAD OF THE**
DEPARTMENT

Mrs .G. Rama Devi   **Mr.R. Lakshmana Rao**
Asst. Professor   **Asst. Professor**

Dept. of CSE   **Dept. of CSE**

EXTERNAL EXAMINER

ACKNOWLEDGEMENT

It is with at most pleasure and excitement we submit our project partial fulfillment of the requirement for the award of Bachelor of Technology.

The project is a result to the cumulate efforts, support, guidance, encouragement and inspiration from many of those for whom we have to give our truthful honor and express gratitude through bringing out this project at the outset as per our knowledge.

I convey my special thanks to our project **Guide Mrs. G. Rama Devi(Asst. Professor)** who has guided, encouraged and tremendously supported me to enhance my knowledge with present working of this project to bring out enriching the quality of project.

I express my appreciativeness to **Mr. R.LAKSHMAN RAO(Asst. Professor) and Head of the Department,** who facilitated us to providing the friendly environment which helped to enhance my skills in present project.

I would also like to extend my gratitude to **Dr. K. V. G. KRISHNA MURTHY, Dean-School of Engineering And Technology, Centurion University of Technology and Management** who has helped us to attain all the requirements of the project.

I convey my sincere thanks to **Dr. RAMANA RAO, Ph. D Registrar of Centurion University of Technology and Management** who provided us with an opportunity to take on project work in well-equipped laboratories of Computer Science Department in our college.

At the outset, we thank to **Sri. G.S.N.RAJU**, beloved **Vice Chancellor of Centurion University of Technology and Management** who is the back bone by providing for completion of this project, Thank you sir.

DECLARATION

I hereby declare that the project entitled **"RESEARCH CENTER FOR DATA SCIENCE AND MACHINE LEARNING"** submitted to the fulfillment of award the degree of **B.TECH (CSE)** in **CENTURION UNIVERSITY OF TECHNOLOGY AND MANAGEMENT**, ANDHRA PRADESH.

R SRI VENKAT SAI   **Reg.No:211801370080**

1.INTRODUCTION:

Data science and machine learning are two of the most exciting and rapidly growing fields in modern technology. Data science involves the extraction of valuable insights from large and complex data sets using statistical and computational methods. Machine learning, on the other hand, is a subfield of data science that focuses on building algorithms and models that can learn from data and make predictions or decisions. The rise of big data and the increasing availability of powerful computing resources have led to explosive growth in both data science and machine learning. Today, these fields are used in a wide variety of applications, from financial modeling and fraud detection to natural language processing and computer vision.

1.1 PURPOSE:

The purpose of a Center for Data Science and Machine Learning is to provide a centralized hub for research, education, and collaboration in these fields. Such centers typically bring together faculty, students, and researchers from different departments and disciplines to work together on projects related to data science and machine learning.One of the primary goals of a Center for Data Science and Machine Learning is to advance knowledge in these fields through cutting-edge research. This may involve developing new algorithms, models, and techniques, or applying existing ones to novel problems and data sets. The center may also focus on developing new tools and technologies to support data.

1.2 SCOPE:

The scope of a Center for Data Science and Machine Learning can be quite broad, depending on the goals and objectives of the center. Some potential areas of focus for such a center might include:

- Research
- Education and training.
- Industry partnerships.
- Policy and ethics.
- Usability.
- Security.

1.3 DEFINITIONS, ACROYNM AND ABBREVATION:

Center: A research institution or organization focused on advancing a particular field or discipline. Data Science: A multidisciplinary field that combines statistics, computer science, and domain expertise to extract insights and knowledge from data.Machine Learning: A subset of artificial intelligence that enables computers to learn from data and improve performance on a specific task without being explicitly programmed.

CDML: Center for Data Science and Machine Learning
AI: Artificial Intelligence
NLP: Natural Language Processing
CV: Computer Vision
Bayesian Stats: Bayesian Statistics
RL: Reinforcement Learning
Graphical Models: Models that use graphs to represent the relationships between variables
Big Data Analytics: The process of analyzing large and complex datasets to extract insights and knowledge
XAI: Explainable AI
Causal Inference: The process of determining the causal relationship between variables in a dataset.

1.4 REFERENCES:

- Researchgate.cutm.ac.in
- Research.cutm/rcfordatascience-and-machine-learning//ac.in.

1.5 OVERVIEW:

The Center for Data Science and Machine Learning (CDML) is a research institution that focuses on the development and application of advanced techniques in data science and machine learning.The CDML aims to foster interdisciplinary collaboration among researchers and practitioners in computer science, statistics, mathematics, and other fields, with the goal of advancing the state-of-the-art in data science and machine learning.The center engages in research on a wide range of topics, including:

- Deep learning
- Natural language processing
- Computer vision Bayesian statistics
- Reinforcement learning
- Graphical models
- Big data analytics
- Explainable AI
- Causal inference

In addition to research, the CDML also offers educational programs, such as courses, workshops, and internships, to help train the next generation of data scientists and machine learning experts.

2.OVERALL DESCRIPTION:

The Center for Data Science and Machine Learning (CDML) is a research institution that focuses on advancing the field of data science and machine learning through

interdisciplinary research and education. The center engages in research on a variety of topics, including deep learning, natural language processing, computer vision, Bayesian statistics, reinforcement learning, graphical models, big data analytics, explainable AI, and causal inference.The CDML also offers educational programs, such as courses, workshops, and internships, to train the next generation of data scientists and machine learning experts. The center collaborates with industry and government organizations to apply advanced techniques in real-world settings and solve complex problems.Overall, the CDML serves as a hub for cutting-edge research, education, and innovation in the field of data science and machine learning.

2.1 PRODUCT PERSPECTIVE:

From a product perspective, the Center for Data Science and Machine Learning (CDML) could be seen as a hub for the development and application of advanced techniques in data science and machine learning. The center's primary products could be:Research outputs: The CDML produces cutting-edge research outputs such as scientific papers, technical reports, and patents that contribute to the advancement of the field of data science and machine learning.

2.2 SYSTEM INTERFACE:

The system interface of the Center for Data Science and Machine Learning (CDML) involves several components that interact with each other to achieve the center's objectives. These components could include:

- Research teams
- Educational programs.
- Industry and government partnerships.
- Tools and platforms.

2.3 USER INTERFACE:

The user interface of the Center for Data Science and Machine Learning (CDML) involves several components that interact with its users, including:

1. Researchers.
2. Students.
3. Industry and government partners.
4. General public.

2.4 HARDWARE INTERFACE:

The hardware interface of a center for data science and machine learning can vary depending on the specific needs and goals of the center. However, there are some common hardware components and interfaces that may be present in such a center:

- Servers.
- Storage.
- Networking.
- GPU
- Workstations
- Visualization systems.

2.5 SOFTWARE INTERFACE:

The software interface of a center for data science and machine learning would depend on the specific research goals and needs of the center. However, there are some common software components and interfaces that are likely to be present in such a center:

I. **Data management software**: To manage large volumes of data, a center for data science and machine learning would need data management software such as databases, data lakes, and data warehouses. These tools help to organize, store, and retrieve data efficiently.
II. **Machine learning and data analysis software**: To perform machine learning and data analysis tasks, the center would need specialized software packages such as Python's scikit-learn, TensorFlow, PyTorch, and R. These software packages provide a range of machine learning and data analysis algorithms and libraries.
III. **Version control software**: Version control software such as Git is used to track changes in code and data, and allows researchers to collaborate effectively on code development projects.

2.6 COMMUNICATION INTERFACE:

The communication interface of a center for data science and machine learning is important for collaboration and knowledge sharing among researchers, as well as for outreach to industry partners and the wider research community. Some common communication interfaces that may be present in such a center. Overall, the communication interface of a center for data science and machine learning should support effective collaboration and knowledge sharing among researchers within the center and with external partners, as well as outreach and promotion of research findings to the wider research community.
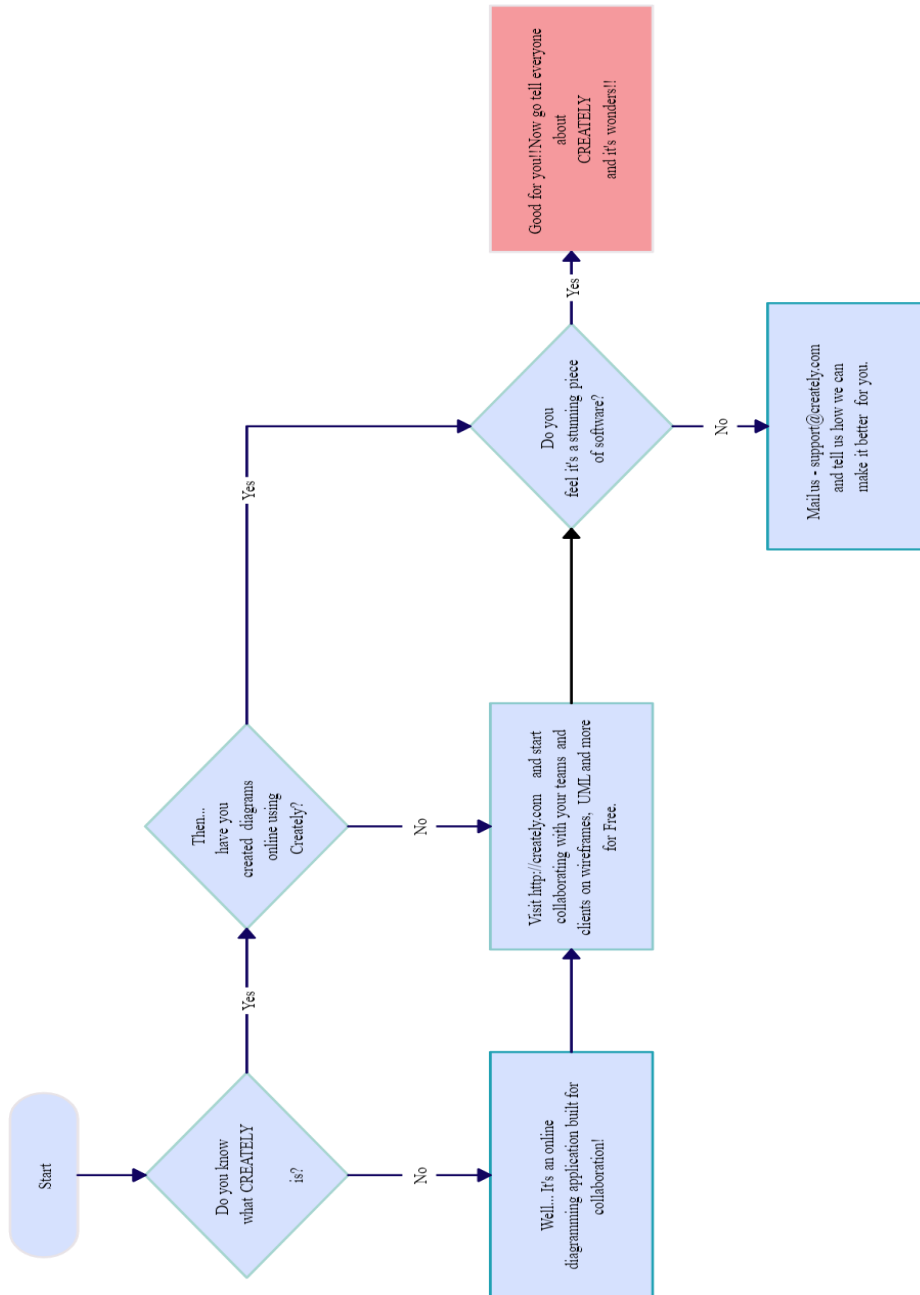
2.7 MEMORY CONSTRAINTS:

Memory constraints can be a major challenge in a center for data science and machine learning, particularly when working with large datasets or complex models that require significant amounts of memory. Some common memory constraints that may be encountered in such a center. Overall, managing memory constraints is an important consideration in a center for data science and machine learning, and requires careful planning and optimization to ensure efficient use of available resources.

2.8 OPERATIONS:

The website executes a number of duties and procedures to promote user interaction and cooperation. A website for communication and technology could include the following important functions:

- Research.
- Education and training.
- Consulting and services.
- Collaboration and partnerships.
- Outreach and communication.

3. FLOWCHART

Start

Do you know what CREATELY is?

No → Well... It's an online diagramming application built for collaboration!

Yes → Then... have you created diagrams online using Creately?

No → Visit http://creately.com and start collaborating with your teams and clients on wireframes, UML and more for Free.

Yes → Do you feel it's a stunning piece of software?

No → Mail us - support@creately.com and tell us how we can make it better for you.

Yes → Good for you!!Now go tell everyone about CREATELY and it's wonders!!

4. USE CASE DESCRIPTION / INTRODUCTION:

A center for data science and machine learning can be a valuable resource for a variety of organizations and industries. Here is a use case description of how such a center might operate for a hypothetical company in the healthcare industry:

4.1 USEFUL FOR A DEPARTMENTS :

The healthcare company is looking to improve patient outcomes and reduce costs by leveraging data science and machine learning. They partner with a center for data science and machine learning to develop and implement a machine learning system that can predict patient readmissions and identify patients who are at risk of developing complications.

4.2 USEFUL FOR A TRAINING:

The center's team of data scientists work closely with the healthcare company's clinicians to identify relevant data sources and develop predictive models. They also provide training to the company's staff on how to use and interpret the machine learning system. Through this partnership with the center for data science and machine learning, the healthcare company is able to leverage the latest advances in machine learning to improve patient care and achieve their business goals.

5. PROJECTS:

The specific projects undertaken by a center for data science and machine learning would depend on various factors such as the center's focus areas, resources, expertise, and collaborations. However, here are some general examples of projects that a center for data science and machine learning might undertake:

- **Developing and improving machine learning algorithms:** This could involve working on new models for specific applications, optimizing existing models, or developing new techniques to improve model accuracy, speed, or interpretability.
- **Building predictive models:** This could involve using machine learning algorithms to analyze large datasets and build models to predict outcomes in fields such as healthcare, finance, or weather forecasting.
- **Natural language processing:** This involves developing algorithms that can understand and process human language, which can be used for applications such as chatbots, sentiment analysis, and machine translation.
- **Computer vision:** This involves developing algorithms that can process visual information, which can be used for applications such as object recognition, face detection, and autonomous vehicles.
- **Data visualization:** This involves developing tools and techniques to visualize large and complex datasets, which can help users understand patterns and relationships in the data.

- **Data mining and pattern recognition:** This involves developing algorithms that can identify patterns and relationships in large and complex datasets, which can be used for applications such as fraud detection and anomaly detection.
- **Social media analysis:** This involves using machine learning algorithms to analyze social media data and gain insights into consumer behavior, brand sentiment, and other trends.

6. LABS:

The CDML has several labs that specialize in different areas of machine learning and data science. Here are a few examples:

- **Computer Vision Lab:** This lab focuses on developing algorithms and techniques for computer vision, including image recognition, object detection, and image segmentation. Researchers in this lab are working on applications such as autonomous vehicles, surveillance, and medical imaging.
- **Natural Language Processing Lab:** This lab specializes in developing algorithms and tools for analyzing and processing natural language data. Research areas include sentiment analysis, language modeling, and machine translation. Applications include chatbots, voice assistants, and text analytics.
- **Deep Learning Lab:** This lab focuses on developing deep learning algorithms and architectures, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and generative adversarial networks (GANs). Researchers in this lab are working on applications such as image and speech recognition, natural language processing, and autonomous systems.
- **Data Mining Lab:** This lab focuses on developing algorithms and techniques for discovering patterns and knowledge from large datasets. Research areas include clustering, association analysis, and outlier detection. Applications include fraud detection, market segmentation, and recommendation systems.

7. LOGIN:

By entering a set of credentials that authenticate their identity, a user can access a computer system, website, or application through the login procedure. A login often entails inputting a username and password. A login is used to make sure that only authorised users may access sensitive data or carry out certain tasks inside a system. A login aids in guarding against security breaches and preventin unauthorised access by asking users to input their credentials. For access to the services offered by many websites and applications, users frequently need to create a login.

8. SEARCH:

Instead of utilising hierarchical links, the system must be able to traverse sites using search capability. Academic portals can demand too many clicks and might be challenging to utilise effectively. Students become frustrated wheNfundamental tasks, like turning in an assignment, require too many steps. Students can easily locate what they're seeking for thanks to a search tool and a hierarchical approach to the academic portal's search functionality.

9. USER CHARACTERISTIC:

9.1 STUDENT:

Students who are interested in pursuing a degree or career in data science and machine learning typically have a strong background in mathematics, computer science, statistics, and programming. Some of the key characteristics of a successful student in a center for data science and machine learning.

9.2 TEACHERS:

The characteristics of a successful teacher in a center for data science and machine learning are somewhat similar to those of a successful student, but with additional skills and experience related to teaching, mentoring, and research. Some of the key characteristics of a successful teacher in a center for data science and machine learning.

9.3 SYSTEM ADMISTRATORS:

The role of a system administrator in a center for data science and machine learning is to ensure that the infrastructure, systems, and applications used by the center are functioning properly and securely. Some of the key responsibilities of a system administrator in this context may include:

1. **Setting up and maintaining hardware and software infrastructure**: System administrators are responsible for setting up and maintaining the servers, databases, networking equipment, and other infrastructure that supports the center's data science and machine learning activities.
2. **Monitoring and troubleshooting system issues:** System administrators monitor the performance of the center's systems and applications, and troubleshoot any issues that arise.
3. **Managing security**: System administrators are responsible for ensuring that the center's systems and data are secure, and implementing security measures such as firewalls, access controls, and encryption.
4. **Supporting users:** System administrators provide technical support to users of the center's systems and applications, and help them with any issues or questions they may have.
5. **Planning for future growth**: System administrators work with the center's leadership team to plan for future growth and ensure that the center's infrastructure is scalable and adaptable to new technologies and research directions.
6. **Implementing backup and recovery procedures**: System administrators are responsible for implementing backup and recovery procedures to ensure that data can be recovered in the event of a system failure or data loss.

These are just a few of the key responsibilities of a system administrator in a center for data science and machine learning.

10. CONSTRAINTS:

10.1 CONSTRAINTS WITH USER INTERFACE:

This system is fairly easy to use and straightforward. All of the system's functionality should be clear to a user with a working knowledge of basic browser navigation.

10.2 HARDWARE RESTRICTIONS:

The system should function on the majority of household desktop and laptop machines that support HTML5 and JavaScript.

10.3 SOFTWARE RESTRICTION:

Firefox 4 and later, Google Chrome 10 and later, and Internet Explorer 8 and later are all recommended for use with the system.

10.4 CONSTRAINTS ON DATA MANAGEMENT:

System must be able to communicate with other components in accordance with their requirements.

10.5 OPERATIONAL RESTRAINTS:

The system's operating server places a cap on the number of concurrent users it can handle.

10.6 SITE ADAPTATION RESTRICTION:

After the system is created, the component will be modified to work with the overall system

10.7 ASSUMPTIONS AND DEPENDENCIES:

The majority of academic portals include a tonne of unnecessary

Features that are never utilized during a class. In addition to introducing certain new

Features that other portal lack, our new system focuses on the characteristics that

Academic institute consumers value the most.

11. SPECIFIC REQUIREMENTS:

11.1 EXTERNAL INTERFACE:

11.1.1: WEB SERVER:

• The web server chosen is Apache:

• Using HTML forms, the user submits data to the web server

- The web server runs PHP as a module, and if the post data is accessible, the PHPscript obtains it.
- The PHP script provides data back to the web server.
- The end-user sees an HTML page as a result from the web server.

11.1.2: PHPAPPLICATION:

PHP was used to create the actual programme that will carry out the procedures. A database will be used to store all the data.

11.1.3: MYSQL DATABASE:

It's an open-source SQL database to store all data which communicates with the

application on the server.

11.2 PERFORMANCE REQUIREMENTS:

Performance requirements are a set of criteria or specifications that specify the speed, capacity, and efficiency with which a system or application must operate. These specifications, which are frequently established by users or stakeholders of the system or application, are used to assess the system's performance and make sure that it satisfies its users' needs.

11.2.1 LOGICAL DATABASE SPECIFICATIONS:

All information, with the exception of files that are stored on the disc, will be saved in the database, including user accounts and profiles, discussion data, messages, etc. A solid database architecture is necessary for the database to support concurrent access and maintain consistency at all times.

11.2.2 DESIGN CONSTRAINTS:

1. SQL will be used for all communication between the portal programme and the database.

2. HTML/CSS will be used to create the portal layout.

3. PHP will be used to create the product.

4. The output needs to be W3C XHTML 1.0 compliant.

5. The source code must adhere to PHP's coding standards.

6. Complete documentation must be available to system administrators .

11.3 SOFTWARE SYSTEM CHARACTERISTICS:

The components of the software are as follows:

1.the PHP program and

2.the Apache web server

3. MySQL, the database

11.3.1 RELIABILITY:

The dependability of the individual components affects the program's overall dependability.

11.3.2 AVAILABILITY:

The system should be available at all times, meaning the user can access it using a web browser, only restricted by the down time of the server on which the system runs. In case of a of a hardware failure or database corruption, a replacement page will be shown. Also in case of a hardware failure or database corruption, backups of the database should be retrieved with the MySQL server and saved by the administrator.

11.3.3 SECURITY:

1. Passwords will be saved encrypted in the database in order to ensure the user's privacy.

2. The user's IP will be logged.

3. The system will be protected against vulnerabilities such as SQL injection attacks.

11.3.4 MAINTAINABILITY:

MySQL is used for maintaining the database and the Apache server takes care of

the site. In case of a failure, a re-initialization of the program is recommended.

11.3.5 PORTABILITY:

The application is Linux-based and should be compatible with other systems. Apache, PHP and MySQL programs are practically independent of the OS-system which they communicate with. The end-user part is fully portable and any system using any web browser should be able to use the features of the application.