Planning Report

1. **Introduction:**

My project's main goal is to create a patient heart failure/attack prediction system by utilising predictive models and cutting-edge machine learning techniques. The aim of this study is to determine whether it is possible to anticipate the risk of a heart attack in individuals.

* 1. **Explanation**

This project's main objective is to develop and deploy a patient heart failure/attack prediction system that makes use of predictive modelling and machine learning methods. This project's main goal is to determine whether it is possible to predict heart attacks, which would allow for early intervention and preventative therapy.

The research will make use of a dataset that includes relevant patient data and medical markers to do this. This dataset will be analysed and learned from using machine learning algorithms, which will produce predictive models that can recognise trends and risk factors related to heart attacks. Subsequently, these models will undergo extensive assessment and validation procedures to guarantee their accuracy and dependability.

Essentially, the goal of this project is to advance the area of healthcare by creating an advanced technology that could help doctors forecast heart attacks in patients, which could ultimately result in better patient care and possibly even save lives.

* 1. **Background**

It is important to acknowledge an understanding of the elements that lead to the occurrence of heart attacks before beginning this endeavour. Numerous studies and an abundance of published works explore the complex nature of heart attacks, each offering insightful information about the different routes and risk factors that contribute to this saddening cardiovascular disaster.

The abundance of available information underscores the complexity of heart attacks, which can result from a multitude of processes. These processes encompass genetic, behavioural, and physiological factors, alongside uncontrollable elements like race, sex, and age. Additionally, factors such as smoking, exercise, stress, and diet exert significant influence *(Ashwell, 1997)*. My work attempts to use the understanding that there are several routes leading to a heart attack to develop a strong system that can recognise various risk factors. Drawing from a thorough understanding of this important subject, my study aims to contribute to the proactive management of heart health by utilising machine learning and predictive modelling approaches. This is not only a stand-alone endeavour; rather, it expands on the body of information found in the literature to provide a fresh and helpful answer.

* 1. **Problems**

The development of the Patient Heart Failure/Attack Prediction System prompts a critical examination of pertinent issues that necessitate in-depth scrutiny. While understanding these limitations is paramount, it is equally crucial to assess potential vulnerabilities within the system, with a view to safeguarding users from harm and unwarranted complications.

A collection of challenges can be outlined as follows:

1. **Ethical Issues:** Designing a system of this nature brings up several ethical issues, mostly related to patient data use, informed permission, data privacy, and possible algorithmic bias. A crucial concern is making sure the system functions morally and respects patients' rights and privacy.
2. **Patient Data Collection:** Gathering thorough and precise patient data is a major task. The availability of significant medical records and data may be restricted, and the calibre of the data may differ greatly. The system's success depends on ensuring the collection of high-quality, standardised data.
3. **Predicting False Positives and Negatives:** Issues with the system's predictions containing both false positives and false negatives. Serious repercussions may arise from false negatives, in which the system fails to recognise a patient who is in danger of having a heart attack. In addition, a high false positive rate can lead to needless anxiety and medical visits. It is crucial to strike the correct balance between specificity and sensitivity. This harmony ensures patient safety and the legitimacy of the system by reducing false negatives while avoiding an excessive input of false positives.
   1. **Context**

Technological advancements have become a driving force in contemporary healthcare, reshaping the landscape of medicine and patient care *(Stoeva, 2020)*. The indispensable role of physical and engineering sciences in medicine is evident as they continually contribute to the development of advanced healthcare techniques, procedures, and equipment. This ongoing process is pivotal in the ever-changing healthcare ecosystem, where machine learning and data-driven solutions are at the forefront.

During this transformative journey, the healthcare industry is undergoing a paradigm shift. It increasingly relies on data-driven solutions to enhance patient care and improve outcomes. This shift toward data-driven healthcare is where my project steps in, poised to make a significant contribution to the global healthcare imperative of proactive healthcare management.

My project focuses on the prediction of heart failure/attack, a critical endeavour given the challenges posed by an aging population and the growing prevalence of chronic conditions. Anticipating and preventing heart-related ailments has never been more essential, and this is where the power of technology, particularly machine learning, shines. By harnessing the capabilities of data analysis and predictive modelling, I can identify at-risk individuals and intervene before these life-threatening events occur.

Furthermore, my project serves as a vital link between cutting-edge technology and today's most pressing healthcare challenges. It's a testament to the synergy between innovation in physical and engineering sciences and the ever-evolving field of medicine. Through this endeavour, I can bridge the gap between the advancements in technology and the critical needs of our healthcare system, striving to provide more effective, efficient, and patient-centred care. As machine learning continues to transform the healthcare industry, this project is a shining example of how I can harness the power of data and technology to improve patient outcomes and meet the healthcare demands of the 21st century.

* 1. **Implications**

As I commence this project, I am acutely aware of the profound implications it will have for both healthcare and patients. The successful implementation of my patient heart failure/attack prediction system is more than just a research project; it has the potential to change the very fabric of healthcare. With this project, I hope to provide healthcare professionals with a powerful tool for early intervention and preventive care, thereby reducing the suffering and financial burden associated with sudden cardiac events. The system promises patients a newfound sense of control over their health, allowing them to make informed decisions and take proactive steps towards a healthier future.

Furthermore, the ethical data handling practises that I uphold in this project set a high standard, emphasising the importance of patient privacy and rights *(Data Protection Act 2018)*. In the grand scheme of things, I see my project seamlessly integrating with the global shift towards data-driven precision medicine, resulting in improved patient care, cost savings in healthcare, and possibly even lives saved. As someone who is deeply invested in the healthcare industry, I see this project as an opportunity to be a change agent, ushering in a new era of early prevention and patient-centred care.

1. **Aims and Objectives**

Given this project's expansive scope, many aims and objectives could be outlined, each contributing to its overarching goals. While some of these aims and objectives may hold strategic value in the long term, it is essential to prioritize those that align most closely with the current state of development and feasibility. Consequently, the following list represents the most significant and attainable goals and objectives that merit my immediate focus, ensuring a pragmatic and effective approach to the project's objectives.

The aims that I will be directing my focus towards encompass:

1. **Developing a Robust Prediction Model:** The aim is to create a machine learning-based prediction model that can accurately identify if an individual is of risk of heart failure or heart attacks in patients.
2. **Ensuring Ethical Data Usage:** Implementation of ethical data handling practices to respect patient privacy and consent while ensuring the system’s compliance.
3. **Enhancing Early Intervention:** Enabling early intervention and preventive care by predicting heart attack risks in a timely and reliable manner, potentially saving lives, and reducing healthcare costs.
4. **Minimising False Positives and Negatives:** Contribute to the field of healthcare by advancing predictive technology that can support medical professionals in providing more effective patient care.
5. **Real-time Data Integration:** Develop a mechanism to collect and integrate real-time data from smartwatch devices, including heart rate, activity levels, sleep patterns, and other relevant physiological metrics.

My targeted objectives encompass:

1. **Model Development:** Develop machine learning algorithms and predictive models capable of analysing the dataset to recognise patterns and risk factors related to heart attacks.
2. **Data Collection and Preprocessing:** Collect, clean, and preprocess a diverse dataset that includes relevant patient data and medical markers for training the prediction model.
3. **Model Evaluation:** Rigorously assess and validate the predictive models to ensure their accuracy, reliability, and generalizability to different patient populations.
4. **User Education:** Educate healthcare professionals and patients about the system's capabilities, limitations, and its role as a predictive tool to manage expectations and promote informed usage.
5. **Continuous Monitoring and Updates:** Establish a mechanism for continuous monitoring of patient data and model performance, allowing for updates and adjustments as medical knowledge evolves.
6. **Tasks and Deliverables**

**Tasks and Deliverables:**

The tasks and deliverables needed to make sure I can complete this project to a professional standard and display the correct output are listed below.

1. **Data Collection and Preprocessing:**

**Task:** Gather relevant patient data, medical markers, and real-time smartwatch data.

**Deliverable:** A cleaned and formatted dataset ready for analysis and model training.

1. **Model Development and Training:**

**Task:** Develop machine learning algorithms and predictive models.

**Deliverable:** Trained prediction model capable of analysing patient data for heart failure/attack risk.

1. **Model Evaluation and Validation:**

**Task:** Rigorously assess and validate the predictive models.

**Deliverable:** A validated model with established accuracy, reliability, and generalizability.

1. **Ethical and Regulatory Compliance:**

**Task:** Establish and implement ethical data handling practices and ensure regulatory compliance.

**Deliverable:** A system that respects patient privacy, complies with relevant regulations, and adheres to ethical guidelines.

1. **User Education and Interface Development:**

**Task:** Create an intuitive and user-friendly interface for patients and healthcare professionals.

**Deliverable:** User education materials and a user-friendly interface to ensure effective system usage.

1. **Continuous Monitoring and Updates Implementation:**

**Task:** Develop a mechanism for continuous monitoring of patient data and model performance.

**Deliverable:** A system that can adapt to evolving health patterns and provide up-to-date risk assessments.

**Scope Considerations:**

The development and implementation of a patient heart failure/attack prediction system falls under the main purview of the project. Despite being a comprehensive project, owing to time and resource constraints, some tasks and deliverables might be deemed out of scope. A noteworthy possibility is that the system's scalability and long-term sustainability will go beyond the current project's parameters.

**Main Project Milestones:**

1. **Milestone for Data Collection and Preprocessing:** Finalisation of data gathering and preparation for modelling and analysis.
2. **Milestone for Model Development and Training:** The heart failure/attack prediction model was developed and trained successfully.
3. **Assessment and Validation of the Model Milestone:** The accuracy and dependability of the model are validated and verified.
4. **Milestone for Ethical and Regulatory Compliance:** Putting in place moral and legal guidelines for managing data.
5. **Milestone for User Education and Interface:** Creating instructional materials and user-friendly interfaces.
6. **Constant Watching and Updates Milestone:** Putting in place a system for continuous data monitoring and system upgrades.

These targets represent significant accomplishments along the way towards implementing the Patient Heart Failure/Attack Prediction System and act as important checkpoints throughout the project.

1. **Gantt Chart**

Using a Gantt chart, I have outlined the key milestones and tasks of my project's timeline. From the start of the project to its completion, this chart will serve as a roadmap.

In the Gantt chart, not only are the major components of the "Final Year Project" included, such as project initiation and key development phases, but also the timeframes for other courses and responsibilities. As a result, I can effectively manage my time and resources, ensuring that I meet all my academic and project-related obligations.

**A screenshot of a computer

Description automatically generated**

1. **Resources**

**Resources for Project:**

In the following section, I highlight the critical resources required for the project's successful completion. These resources include both data and knowledge, and they will serve as the foundation for the project's outcomes.

1. **Research Studies:** It is critical to have access to important medical studies and research on heart attack predictions, prevention, and diagnosis. These academic articles bring vital insights into the field of cardiology and predictive healthcare to the project.
2. **Heart Health Data:** The foundation of this initiative is extensive data on heart attack patients, including clinical history, test findings, and medical records. This data is essential to achieving the project's goals and serves as the foundation for creating predictive models.
3. **Biological Knowledge:** It is crucial to have an in-depth understanding of the physiology, anatomy, and biochemical causes of heart attacks. The inner workings of the heart must be understood to develop a trustworthy prediction system.
4. **Predictive Models:** Identify appropriate predictive models through research for the kind of prediction needed for the project. A crucial first step in accomplishing the intended project goals is choosing the appropriate models.
5. **Smartwatch Data:** The use of smartwatches for the prediction of heart attacks and the possibility of collecting data is a contemporary resource. In addition to providing a novel method for collecting real-time health data, smartwatches may significantly improve the accuracy of predictions.
6. **Software and Tools:** The project incorporates a wide range of software and tools, including Python for data analysis and modelling, Pandas for data manipulation, Scikit-Learn for machine learning algorithms such as logistic regression, Pyspark for big data processing, MySQL for structured data storage, Ruby for iOS health data smartwatch configuration, and AWS for cloud-based infrastructure, all of which contribute to the project's data processing and analysis capabilities.

**Sources of Information:**

1. **Academic Literature:** Academic literature serves as a wealth of information and includes research papers, scientific publications, and medical databases. It guarantees that the work is based on well-established research and serves as the project's intellectual framework.
2. **Healthcare Professionals:** It is necessary to consult cardiologists and other medical experts. Their in-depth clinical knowledge and experience offer a practical viewpoint on heart health and prognosis.
3. **Online Communities:** Engaging in active engagement in virtual communities, forums, and discussion platforms links the project to an abundance of shared information. These groups are a great resource for learning about the newest techniques and trends.
4. **Machine Learning Experts:** The project's technical skills are strengthened by consulting with experts in predictive modelling and machine learning approaches. Predictive model selection and execution are guided by their experience.
5. **Smartwatch Manufacturers:** Speaking with technical specialists or smartwatch makers can provide me with information about what wearable technology is capable of. The project's success depends on recognising the potential of smartwatches in data collecting.
6. **Publicly Available Datasets:** The project's robust data foundation is built on publicly available datasets related to heart health and smartwatch data. In addition to supplementing the project's resources, these datasets facilitate comparisons and benchmarking.
7. **Risks**

Any project, including this one, is vulnerable to several risks that could jeopardise its success. Some of these dangers are unexpected, while others are predictable. In this section, I will identify potential risks, explain why they exist, and suggest mitigation techniques to ensure the project's success.

|  |  |  |
| --- | --- | --- |
| **Risk** | **Cause** | **Proposed Solution** |
| Data Quality Issues | Inaccurate or incomplete data can result in false predictions. | Implement precise data validation and cleanup procedures. Use data from reliable sources. |
| Technical Challenges | Building complex machine learning models can be technically challenging, demanding a great deal of expertise. | Maintain a diverse professional network for consultation and conduct regular code reviews. |
| Ethical and Privacy Concerns | Mishandling of patient data may result in ethical and legal issues. | Maintain strict compliance with data protection regulations. De-identify or conceal sensitive information or Obtain participants' informed consent. |
| Model Overfitting | Models that have been overly tuned to training data may perform poorly on new data. | Implement cross-validation techniques and regularization methods to prevent overfitting. |
| Resource Constraints | Time constraints may impact project deliverables. | Prioritise tasks and use resources wisely. |
| Changing Project Scope | Evolving project requirements can lead to scope creep. | Maintain a clear and documented project scope. |
| Lack of User Acceptance | Patients and healthcare professionals may be hesitant to use the prediction system. | Include end users from the beginning. Conduct user testing and feedback gathering to improve the system. Users should be trained and supported. |
| Regulatory Changes | New healthcare regulations could have an impact on data handling and model deployment. | Stay updated with relevant regulations and adapt the project to comply with any changes. |
| External Dependencies | Using third-party tools or data sources may result in disruptions. | Determine critical dependencies early on and have backup options or contingency plans in place in the event of a disruption. |

1. **Career Aspirations**

Looking ahead to my career goals, this project holds a lot of promise in terms of preparing me for a role in data engineering, particularly in the healthcare sector. It offers a once-in-a-lifetime opportunity to learn essential tech-stacks such as Pyspark, AWS, Python, Ruby, and various databases, all of which are highly sought after in the data engineering field *(Nudurupati, 2021)*.

The project's emphasis on big data processing with Pyspark and cloud computing with AWS aligns perfectly with the growing demand for professionals who can handle and analyse large datasets in a cloud environment. Furthermore, this project's hands-on experience managing and integrating data from various sources is directly transferable to real-world data engineering challenges. Importantly, the project allows participants to delve into the complexities of healthcare data, a critical domain in the healthcare industry.

Furthermore, practical experience with database management, the development of predictive models using machine learning, and the training of these models on collected data enhances my expertise. As I work through the complexities of this project, I am not only laying a solid foundation in data engineering but also honing problem-solving and analytical skills, both of which are essential in the field. The inherent project management elements of the project, such as task prioritisation, time management, and milestone tracking, also contribute to my professional readiness.

By highlighting these skills and experiences on my CV, I hope to position myself as an appealing candidate for data engineering roles, particularly those in the healthcare sector, where a combination of technical prowess, domain knowledge, and predictive modelling using machine learning is in high demand. With its diverse and practical tech stack exposure and predictive model development, this project will undeniably enrich my CV and open doors to a fulfilling and impactful career in data engineering.

1. **Legal, Social, Ethical and Professional Issues (LSEPIs)**

The Legal, Social, Ethical, and Professional Issues (LSEPIs) are essential for the direction of my project. These factors have a significant impact on both the project's outcomes and its execution process as I embark on developing a prediction system for patient heart failure/attacks. Legal standards, data protection rules, and societal expectations must all be met. Safeguarding, health and safety, and risk evaluations are all examples of ethical considerations. Maintaining professional standards through ethical coding practises, model verification and honest result sharing is critical to the integrity of my project.

**LSEPIs Relating to Project Outcomes:**

1. **Legal Issues:**

**Project Outcome:** A prediction system for patient heart failure/attacks is being developed through a project that involves the application of predictive modelling and machine learning techniques, with the aim of addressing legal issues.

**Link:** Predictive modelling of patient data is heavily reliant on legal requirements, including data protection legislation. The outcome of the project is to create a system that ensures data security and privacy while also complying with legal regulations. Legal standards, particularly data protection regulations such as the Data Protection Act 2018, are critical in maintaining the project's data security and privacy (Data Protection Act 2018, c 12).

1. **Social Issues:**

**Project Outcome:** Although the exact project outcome with respect to social issues requires additional explanation, proactive health management can have a significant impact on society.

**Link:** Participating with healthcare experts and patients throughout the project guarantees that it is in line with society's demands and ideals. We want the initiative to be relevant, inclusive, and valuable to the community.

1. **Ethical Issues:**

**Project Outcome:** The project's goal is to build a system that respects patients' rights and privacy, and ethical data handling practises are critical to that goal.

**Link:** Ethical considerations are essential to the project's outcomes, as they ensure that the project maintains high ethical standards in data management and adheres to ethical principles throughout the development process.

1. **Professional Issues:**

**Project Outcome:** By adhering to the rules and standards of the profession, the initiative helps to maintain professional competence and integrity.

**Link:** Professional issues, such as ethical coding practices and maintaining transparency, have a direct impact on the project's goal of developing a system that meets the requirements of the industry.

**LSEPIs Relating to the Process of Carrying Out the Project:**

1. **Legal Issues:** The legal element of the project extends to how data is collected, stored, and used throughout. Data protection and privacy rules should be always followed. Clear data usage regulations and permission methods must be in place.
2. **Social Issues:** The project's implementation should promote openness and inclusivity. Involving stakeholders in the creation and validation process, including healthcare professionals and patients, ensures that the project matches with societal needs and values.
3. **Ethical Issues:** Ethical considerations, such as safeguarding, health and safety, and risk assessments, remain critical in the project's implementation. It is a continuous responsibility to ensure that data gathering practises respect patient privacy and rights.
4. **Professional Issues:** Maintaining professional integrity throughout the project's execution is critical. Adhering to ethical coding practises, verifying the accuracy of predictive models, and sharing results publicly and honestly are all part of upholding the norms of the data engineering and healthcare professions.
5. **References**

Ashwell, M. (1997) ‘Introduction’, in *Diet and heart disease: A round table of factors*. London: Chapman & Hall.

Stoeva, M. (2020) Technological advancement as a driving factor of contemporary healthcare, Shibboleth authentication request. Available at: https://link-springer-com.ntu.idm.oclc.org/article/10.1007/s12553-020-00434-5 (Accessed: 26 October 2023).

Data Protection Act 2018 c 12. 2018. Available at: https://www.legislation.gov.uk/ukpga/2018/12/contents/enacted (Accessed: 4 April 2023).

Nudurupati, S. (2021) Essential Pyspark for Scalable Data Analytics, O’Reilly Online Learning. Available at: https://learning.oreilly.com/library/view/essential-pyspark-for/9781800568877/?sso\_link=yes&amp;sso\_link\_from=nottingham-trent-university (Accessed: 26 October 2023).