VAAL UNIVERSITY
OF TECHNOLOGY

Inspiring thought. Shaping talent.

FACULTY:	Applied and Computer Sciences
DEPARTMENT:	Computer Sciences
SUBJECT:	Business Analysis 3.2
SUBJECT CODE:	AIBUY3A
LECTURER:	Mr. Matsela
MODERATOR:	Mr. XP Piyose
DUE DATE:	13 October 2025

DECLARATION

We, **Mantso Mbali Whitney, Mintirho Mahani, Ntokoto Love Mkansi, Nyathi Mbuyelo Emotion, Manganyi Tiyani Harmony, Mahlasela Liffy, Mcacisi Siyabonga Sithole, Kgaasi Koketso, Khoza Katekile Relation,**

declare that the contents of this project represent our own unaided work, and that the project has not previously been submitted for academic examination towards any qualification. Furthermore, it represents our own opinions and not necessarily those of the Vaal University of Technology.

Signed	**Date**
Signed	**Date**
Signed	**Date**
Signed	**Date**
Signed	**Date**
Signed	**Date**
Signed	**Date**
Signed	**Date**
Signed	**Date**

Table of Contents

Business background

CarePredictAI is a smart patient management system built to ease the pressure on South African healthcare facilities. Instead of forcing everyone to wait in the same long queues before receiving medical attention, the system uses patient symptoms to sort people into clear categories, such as critical (high risk), moderate, or normal (low risk), and uses a chatbot check-in system. Behind the scenes, CarePredictAI operates using Python-based machine learning models, which are trained on medical datasets. Looking ahead, the solution can be scaled to clinics and hospitals nationwide, creating a fairer, faster, and more reliable way to manage patient care.

Problem definition

Many individuals go home without receiving treatment from healthcare facilities. Hospitals and clinics in South African communities still take patients manually, and they face the persistent challenge of long queues and a paper-based system, which delays access to care, overwhelms healthcare staff (misplaces files and makes duplicates), and reduces overall patient satisfaction. Healthcare workers must make rapid decisions with limited data and under high pressure, often relying on experience rather than real-time analytics.

This leads to inconsistent prioritisation, especially during busy hours or emergencies when accurate risk assessment is most critical. Patients with low risk often wait the same amount of time as those with high risk, leading to overcrowding and misallocation of medication. This inefficiency is partly due to the lack of effective classification and prediction tools that can quickly identify high-risk patients who need urgent attention. In hospitals, people must wait to create a file, which can take time before they receive medical attention.

Benefits of AI

By applying AI to patient intake and clinical data, hospitals and clinics can automatically predict which patients are at higher risk of complications or deterioration. This allows healthcare staff to prioritise urgent cases, streamline resource allocation,

and reduce unnecessary waiting times. For the local municipality, the benefits are significant: improved community health outcomes, better use of limited medical staff, and reduced frustration among patients. This AI-driven solution ensures that vulnerable individuals receive timely care while maintaining a smooth patient flow in hospitals and clinics.

AI Solution

CarePredictAI utilises artificial intelligence to help hospitals and clinics manage patients more effectively. When a patient registers, the system gives them a chatbot which will assist them in making a profile. The chatbot system will allow a patient to enter details such as name, age, symptoms, medical history and lifestyle, after which the AI model will estimate their risk level. The chatbot will save all the information that was provided by the patient. This way, the hospital already has the patient's details, reducing waiting lines and helping doctors prepare in advance.

The solution applies machine learning in healthcare, which is a part of the project theme *"AI Solution for Industries".* The model is trained on medical datasets and simulated patient records, and its accuracy is tested to make sure the results are reliable. The chatbot applies natural language processing for patient interaction and ensures that the solution is practical for real-world use. By combining AI risk prediction, CarePredictAI provides a fair and efficient healthcare solution that can be deployed in hospitals and clinics across South Africa.

Business objectives

The primary objective of the AI solution is to predict patient risk levels at the point of entry, enabling hospitals and clinics to classify more effectively, reduce waiting times, and optimise community healthcare delivery. At the same time, we aim to develop a chatbot that asks patients for basic information and symptoms and then provides them with a platform to fill in. In this way, the system supports doctors and administrators in resource planning and decision-making.

Requirements

For our project to work, we need a patient dataset that includes age, symptoms, gender, medical history, and lifestyle because these will be used to train and test the AI model. Hospital staff will also need proper training so that they can use the chatbot and AI system effectively in a real environment. A secure way of storing patient data is required to make sure sensitive information is protected and kept private. On the technical side, we need machine learning models to process the data and make predictions, and we will be using Python as the main programming language since it has powerful libraries and tools for AI development.

Constraints

Our solution has some constraints. The dataset may have missed or imbalanced records. Limited availability of high-quality patient data. Resistance to change from healthcare staff. Infrastructure limitations in rural or underfunded clinics. Ethical concerns regarding bias and fairness in AI predictions. Lastly, government acceptance is also a constraint, because the system can only be used if it follows healthcare regulations and policies, and approval processes may take time or face restrictions.

Risks

We also considered risks that could affect the project. Technical risks include the possibility of the model overfitting or being slow to train on large datasets. Ethical risks come from the possibility that people may misuse predictions without proper doctor oversight, and incorrect predictions may delay critical care if not monitored. Data privacy breaches could compromise patient trust. The cost of implementation may be high for resource-limited hospitals.

Business Success Criteria

The success of CarePredictAI will be measured by how well it helps hospitals and clinics improve their daily operations and service delivery. We will know the project is successful if it meets the following points:

- Reduced Waiting Times: The system should help reduce long queues and patient waiting time by at least 40–60% compared to the normal manual process.

- Better Risk Prediction: The AI model should reach at least 80% accuracy in classifying patients into high-, medium-, or low-risk levels.

- Saving Costs: By automating file creation and sorting patients automatically, hospitals can reduce paper use, overtime pay, and other admin costs by around 30%.

- Revenue Potential: The system can later be licensed or sold to other hospitals or clinics, helping generate income to support maintenance, updates, and new features.

- User Satisfaction: At least 70% of staff and patients should find the Chatbot and system easy to use and helpful for getting faster service.

- Data Protection: All patient data must stay secure and follow South Africa's POPIA law to protect privacy.

- Scalability: The system should be easy to expand to other clinics or hospitals without major technical changes.

If these goals are achieved, CarePredictAI will not only improve healthcare efficiency but also create value and sustainability for hospitals and local communities.

Tools and Techniques

Tools:

- PyCharm: main development environment to build and test the Python code for the AI model and chatbot, debuggers, and edit faster

- GitHub: used for version control and group collaboration
- Python 3.13: works as an interpreter
- XGBoost: it's fast, accurate, and good at handling different types of data. It helped us train the model to predict risk level
- TF-IDF (NLP): used it to help the AI understand symptoms written in normal sentences. It breaks down the words and finds which ones are most important
- Label Encoder(from scikit-learn): used it to turn words like gender, lifestyle, and medical history into numbers so the AI could understand them
- Pandas: used to load and clean data. It helped us organise everything into tables so the AI could learn from it
- CSV logging

Techniques:

- Supervised learning: to train the AI model with a labelled dataset for our chatbot.
- SWOT Analysis: We use this strategic technique to find our strengths and weaknesses.
- Analysing: look for datasets and analyse them

Machine Learning Approach

CarePredictAI uses supervised machine learning to classify patients based on risk levels, which are low, moderate/medium, or low. We used a machine learning method called XGBoost. It looks at the user's age, gender, symptoms, medical history, and lifestyle to predict how serious the situation might be. We trained the model using clean data and tested it to make sure it works well. Then, as time goes on, the AI will improve its predictions using machine learning.

Data

We used clean and relevant data like age, gender, symptoms, medical history, and lifestyle. We removed any missing parts and changed the data into a format the AI can understand. This helped the model learn better and give accurate results. For

testing purposes, simulated or publicly available healthcare datasets (e.g., Hugging Face (Ivimuth) medical datasets) will be used.

Model

The AI model predicts a patient's risk level immediately after chatbot interaction.

How it works

It follows the data life cycle:

1. Data collection
2. Pre-processing
3. Feature extraction
4. Model training
5. Evaluation

The trained model will be stored and occasionally learn to improve its accuracy. The model will output a probability score that determines a patient's priority level in the queue.

Always improving

Our AI is designed to keep getting better. It saves every prediction to a file so we can learn from it later. It also handles new or unexpected answers from users without crashing. This makes it more reliable and ready for future updates.

Time Series Analysis

We didn't use time-based tracking in this version. The AI looks at one-time input only. But in future, we want to add time series so it can follow how a patient's health changes over days or weeks.

Solution Techniques

We used machine learning (XGBoost) to train a model that predicts the risk level of a patient. We cleaned the data, changed words into numbers using TF-IDF for symptoms, and used label encoding for things like gender, lifestyle, and medical history. This helped the AI understand the data better and make accurate predictions.

Natural Language Processing, speech recognition or speech synthesis

We used Natural Language Processing (NLP) to help the AI understand symptoms written by the user. We used a method called TF-IDF, which breaks down the words and finds important patterns. We didn't add voice input or speech output yet, but we plan to include that in future versions.

Deep Learning

We didn't use deep learning in this version. Deep learning uses brain-like models called neural networks. Instead, we used XGBoost, which is a powerful and fast machine learning method. In future, we can try deep learning for more advanced tasks. TensorFlow or Keras will be used to build these models in future versions, improving scalability and accuracy when dealing with large hospital datasets.

Other features

We built a chatbot that talks to the user, asks questions and gathers patient information, and gives a risk level prediction. It's friendly, easy to use, and saves every result to a file. It also works well even if the user types something unexpected. This feature not only speeds up the intake process but also ensures that hospitals have real-time digital records of all visitors, improving operational management and patient satisfaction.

Conclusion

Our project shows how AI can help solve real problems in the healthcare industry. By predicting patient risk levels and reducing waiting lines through a chatbot check-in system, we can improve hospital efficiency and save lives. Even though the project is a prototype with some limitations, it demonstrates how technologies from the fourth industrial revolution can be used to make a real impact on people's lives in South Africa and beyond.