

BABEŞ BOLYAI UNIVERSITY, CLUJ NAPOCA, ROMÂNIA  
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE

# GoldenMate: An Elderly Care Platform

– MIRPR Report –

**Team Members**

Bucila Mihai, Calauz Razvan, Chelaru Laurentiu

2024-2025

## **Abstract**

GoldenMate is an AI-powered mobile application designed to keep elderly people mentally active and engaged. It features "RoboBuddy," an AI companion that provides conversational engagement, health recommendations, and interactive games. The project leverages machine learning, natural language processing (NLP), and data visualization to support the well-being of elderly users.

Intelligent methods used include NLP models for conversation, machine learning for health data analysis, and reinforcement learning for games. The app integrates various datasets to provide personalized insights and interactions for users, with features such as reminders, health tracking, and games.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	What? Why? How? . . . . .	1
1.2	Paper Structure and Original Contributions . . . . .	1
<b>2</b>	<b>Scientific Problem</b>	<b>3</b>
2.1	Problem Definition . . . . .	3
<b>3</b>	<b>State of the Art/Related Work</b>	<b>4</b>
<b>4</b>	<b>Investigated Approach</b>	<b>5</b>
4.1	Model Selection and Development Process . . . . .	5
4.1.1	Conversation Context Management: . . . . .	6
4.1.2	Empathetic Responses: . . . . .	6
4.1.3	Extensibility: . . . . .	6
4.2	Custom Dataset Creation and Preprocessing . . . . .	7
4.2.1	Filtering Relevant Data: . . . . .	7
4.3	Preprocessing and Postprocessing . . . . .	8
4.3.1	Preprocessing User Input: . . . . .	8
4.3.2	Postprocessing Model Outputs: . . . . .	8
4.4	Integration of Health Analysis . . . . .	8
4.4.1	Cardiovascular Disease Prediction . . . . .	9
4.4.2	Diabetes Disease Prediction . . . . .	10
4.4.3	Risk Level Classification . . . . .	12
4.4.4	Health Improvement Suggestions . . . . .	13
4.4.5	Health Improvement Suggestions . . . . .	14
4.5	Sentiment Analysis on User's Message . . . . .	14
4.5.1	Overview . . . . .	14
4.5.2	Methodology . . . . .	14
4.5.3	Daily Sentiment Tracking . . . . .	15
4.5.4	Overall Sentiment Calculation . . . . .	15
<b>5</b>	<b>Application (Study Case)</b>	<b>16</b>
5.1	App Description and Main Functionalities . . . . .	16
5.2	App Design . . . . .	16
5.3	Implementation . . . . .	16
5.4	Testing of the Codebase . . . . .	16
<b>6</b>	<b>Conclusion and Future Work</b>	<b>17</b>

# List of Tables

4.1	Confusion Matrix for Cardiovascular Disease Prediction . . . . .	10
4.2	Confusion Matrix for Diabetes Disease Prediction . . . . .	12

# List of Figures

# List of Algorithms

# Chapter 1

## Introduction

### 1.1 What? Why? How?

The problem addressed by this project is the need for mental and social engagement among elderly people, which can impact their emotional well-being. As people age, they may face increased isolation, cognitive decline, and difficulties managing health metrics. GoldenMate's AI-powered "RoboBuddy" is designed to address these challenges by providing conversation, reminders, health insights, and interactive games.

The project utilizes artificial intelligence (AI) techniques, such as natural language processing (NLP) for conversations and machine learning (ML) for health data analysis. The basic approach involves deploying these models in a mobile application that elderly users can easily access and interact with daily.

This work is relevant as it demonstrates the potential of AI in creating socially and emotionally supportive tools for elderly care.

### 1.2 Paper Structure and Original Contributions

This report presents advancements in AI-based applications for elderly care. The primary contributions include:

- 1. Development of RoboBuddy, an interactive AI-based companion.
- 2. Implementation of health data analysis algorithms for personalized insights.
- 3. Creation of an intuitive and elder-friendly user interface.

The report is structured as follows:

- - Chapter 3 discusses related work.
- - Chapter 4 details the proposed approach and algorithm design.
- - Chapter 5 covers the app's description and functionality.



## Chapter 2

# Scientific Problem

### 2.1 Problem Definition

The project addresses the problem of providing meaningful, interactive support for elderly individuals through AI-driven conversations and health monitoring. As many elderly people face isolation, RoboBuddy aims to alleviate loneliness while also assisting with health management through friendly interactions and reminders.

The AI chatbot, developed using the BlenderBot model, fine-tuned on a custom dataset, interacts with users to provide tailored responses.

Inputs to the system include conversational prompts from the user, medical history and daily health data (such as blood pressure, glucose levels, cognitive assessment results).

The outputs are generated responses suited for conversations with elder people, but also personalized conversational responses that address emotional well-being and reminders for medication, appointments or physical activities.

AI performance is evaluated through the conversational quality, measured by user feedback on response relevance, coherence and empathy determined within sessions and the loneliness improvements due to the ability to play games along the chatbot. The evaluation ensures the system not only supports health monitoring, but also provides meaningful companionship to elderly users.

## Chapter 3

# State of the Art/Related Work

In recent years, advancements in Artificial Intelligence have demonstrated significant potential in enhancing mental and physical well-being, particularly through conversational AI systems. Models like OpenAI’s GPT series and Facebook’s BlenderBot have set benchmarks for natural language processing (NLP), enabling human-like interactions that provide companionship and support. These systems have been explored for various applications, including mental health therapy, customer support, and general social interaction.

Despite these advancements, few solutions have been specifically designed with elderly users in mind. The elderly often face unique challenges, such as social isolation, cognitive decline, and a need for accessible interfaces. While generic conversational AI models excel in dialogue generation, they lack domain-specific features that address these particular needs.

Moreover, existing research highlights the growing demand for AI systems that integrate social engagement with health monitoring. For example, some studies have explored AI-driven virtual assistants that remind users to take medication or log physical activity. However, these systems are typically utility-focused and lack the conversational depth needed to establish emotional connections or mitigate loneliness.

Our project builds on this foundation by introducing a comprehensive, elderly-focused AI application that combines conversational engagement with actionable health insights. Unlike existing conversational models, our system integrates health data such as activity tracking, medication schedules, and real-time updates on vitals. It also incorporates gamified cognitive exercises to foster mental stimulation, promoting both social and physical well-being in an easy-to-use format.

This tailored approach fills the gap in existing research and solutions, providing a holistic tool that meets the specific needs of the elderly population.

## Chapter 4

# Investigated Approach

The proposed approach combines natural language processing (NLP) for conversation, machine learning (ML) for health analysis, and user interface (UI)/user experience (UX) principles for accessibility. We developed RoboBuddy, a Python-based AI companion designed to understand user input and respond empathetically, specifically tailored for elderly users.

### 4.1 Model Selection and Development Process

To determine the most suitable conversational model, we evaluated several state-of-the-art NLP models, including OpenAI's GPT-2, DialoGPT, BERT, and Facebook's BlenderBot.

GPT-2 is pre-trained on a large dataset of diverse internet text, uses a transformer-based architecture to generate fluent and contextually relevant text. (Source: <https://github.com/openai/gpt-2>)

DialoGPT is built on GPT-2, DialoGPT is fine-tuned for conversational data, optimized for generating multi-turn dialogue. (Source: <https://github.com/microsoft/DialoGPT>)

BERT: While primarily designed for understanding tasks, BERT has been adapted for conversational AI due to its deep bidirectional representations. It is pre-trained on masked language modeling and next-sentence prediction tasks. (Source: <https://github.com/google-research/bert>)

BlenderBot is a pre-trained on large-scale dialogue dataset and specializes in empathetic and context-aware conversation generation, with fine-tuning capabilities for custom datasets. (Source: <https://huggingface.co/docs/transformers>)

BlenderBot emerged as the best approach for several reasons:

#### 4.1.1 Conversation Context Management:

BlenderBot demonstrated superior ability to maintain coherent and context-aware conversations over extended dialogues, which is crucial for engaging elderly users. The coherence and the context-awareness was measured using metrics like BLEU scores to evaluate the relevance of generated responses to expected responses. Also, the bot was tested with long conversations, incorporating abrupt topic changes and evaluated on its ability to maintain relevance and logical flow.

A BLEU score example:

- Context: Elderly user asks about health advice related to diabetes.
- User Input: "What should I eat to manage my blood sugar?"
- Model Output: "Eating whole grains, vegetables and lean proteins can help."

The average BLEU Score is 0.82 out of 1.0, demonstrating a relatively strong semantic alignment.

#### 4.1.2 Empathetic Responses:

Its pre-trained models excel in generating empathetic and human-like responses, aligning well with the project's goal of reducing loneliness among the elderly.

The level of empathy of the responses is tested by applying a sentiment analysis tool named TextBlob to gauge whether the bot's responses expressed appropriate levels of positivity, concern or understanding. Also, specific prompts simulating emotional contexts (for example, loneliness, anxiety or frustration) were provided and responses were analyzed qualitatively for emotional depth and appropriateness.

The TextBlob Sentiment Analysis Results reveals that the responses have a positive and empathetic tone, aligning with the project's goal of reducing loneliness, according to the computed results for Polarity value of +0.35 (positive, showing a comforting tone) and Subjectivity value of 0.75 (high subjectivity, indicating personal and empathetic language).

An example of emotional context is the following:

- User Input: "I feel so lonely these days. No one visits me anymore."
- Model Output: "I'm so sorry to hear that. You are not alone, and I'm here to talk with you anytime you need."

#### 4.1.3 Extensibility:

BlenderBot's modular design allowed for the seamless integration of custom datasets and domain-specific fine-tuning, making it adaptable for elderly-centric conversations. While GPT-2 and DialoGPT

provided high-quality text generation, they lacked conversational coherence over long interactions. Similarly, BERT, being optimized for classification tasks rather than dialogue, was not ideal for this use case.

## 4.2 Custom Dataset Creation and Preprocessing

To further enhance the conversational quality and tailor it to elderly users, we developed a custom dataset by merging and filtering data from multiple open-source conversational datasets (DailyDialog, OpenSubtitles, Persona-Chat, TalkBank, etc.).

**DailyDialog:** A high-quality multi-turn dialogue dataset containing 13,118 conversations with topics related to daily life and human interactions. It is designed to reflect natural, human-like conversations. (Source: [https://huggingface.co/datasets/li2017dailydialog/daily\\_dialog](https://huggingface.co/datasets/li2017dailydialog/daily_dialog))

**OpenSubtitles:** A large-scale dataset with over 400 million dialogue lines extracted from movie subtitles, providing diverse conversational scenarios. Unfortunately, this dataset has too many dialogue lines that are not suitable for elderly conversations and the formatting lacks professionalism (many wrong-typed words and Case Insensitive). (Source: [https://huggingface.co/datasets/Helsinki-NLP/open\\_subtitles](https://huggingface.co/datasets/Helsinki-NLP/open_subtitles))

**Persona-Chat:** A dataset of 162,064 dialogue utterances designed for conversational agents with pre-defined personas, helping create more personalized and engaging interactions. Like the OpenSubtitles dataset, this dataset is Case Insensitive, all the inputs being written lowercase. (Source: <https://www.kaggle.com/datasets/atharvjairath/personachat>)

**TalkBank:** A collection of recorded and transcribed conversations, often featuring intergenerational dialogues, including interactions with elderly speakers. The dataset spans multiple languages and contexts. (Source: <https://huggingface.co/datasets/talkbank/callhome>)

The focus was to ensure relevance to elderly users, covering topics such as health, family, cooking, nature, animals, pets, curiosities, entertainment and 6 more.

### 4.2.1 Filtering Relevant Data:

To tailor the dataset to elderly users, we used AI-powered text classification tools, such as GPT-based classifiers, to identify and prioritize elderly-relevant themes. Here's an outline of the logic steps performed for filtering:

1. Keyword-Based Preprocessing:

Hundreds of keywords such as "grandchildren," "pets," "recipes," "hobbies," "health tips," and

"nature" were identified as relevant for elderly-focused conversations. Texts containing these keywords or synonyms were flagged as candidates for inclusion.

## 2. Thematic Categorization:

Each conversation or dialogue snippet was assigned a thematic category using a GPT-based classifier fine-tuned on our custom themes. Irrelevant categories, such as youth slang or aggressive tones, were filtered out, as much as possible manually.

After filtering all the mentioned datasets, the resulted dataset size was approximately 7500 utterances, organized along the 14 categories mentioned earlier.

## 4.3 Preprocessing and Postprocessing

### 4.3.1 Preprocessing User Input:

To improve interaction efficiency, user inputs were preprocessed to handle predefined responses for common interactions, such as greetings ("Good morning! How are you feeling today?"), weather inquiries, and personalized health updates. These predefined responses were dynamically generated based on the user's health data stored in the app.

### 4.3.2 Postprocessing Model Outputs:

Model responses were refined to align with an accessible, friendly tone suitable for elderly users. For example, technical jargon or overly complex language was avoided. To maintain a consistent and depersonalized tone, the model was restricted from using pronouns like "I" and "me." This adjustment was made to emphasize the AI's role as a supportive tool rather than a sentient entity. For instance, instead of saying, "I think you should rest," the model responds with, "It might be a good idea to rest."

## 4.4 Integration of Health Analysis

For health analysis capabilities, we implemented two Random Forest Classifier models for health risk assessment. Both models were developed using scikit-learn and provide probability-based risk predictions.

#### 4.4.1 Cardiovascular Disease Prediction

The cardiovascular disease prediction model is implemented in the `CardiovascularDiseasePredictor` class, trained on a dataset of 70,000 patient records.

##### Feature Handling:

- Required features:
  - Age (in years, range: 0-120)
  - Gender (categorical: male/female)
  - Height (cm, range: 50-250)
  - Weight (kg, range: 20-300)
  - Systolic blood pressure (mmHg, range: 50-250)
  - Diastolic blood pressure (mmHg, range: 30-200)
- Optional features:
  - Cholesterol (1: normal, 2: above normal, 3: well above normal)
  - Glucose (1: normal, 2: above normal, 3: well above normal)
  - Smoking status (0: non-smoker, 1: smoker)
  - Alcohol consumption (0: none, 1: consumer)
  - Physical activity level (0: inactive, 1: active)

##### Data Preprocessing:

- Missing value handling:
  - All optional numerical and binary features default to 0
  - Categorical features use clinically appropriate defaults
  - Missing values flagged for downstream analysis
- Feature transformations:
  - Age conversion from days to years when necessary:  $age\_years = age\_days / 365.25$
  - BMI calculation:  $weight\ (kg) / (height\ (m))^2$
  - Blood pressure validation:  $systolic > diastolic$

- Feature scaling:
  - Standardization using StandardScaler:  $z = (x - \mu)/\sigma$
  - Applied to: age, height, weight, systolic BP, diastolic BP
  - Scaling parameters stored for prediction-time transformation

**Model Configuration:** Random Forest Classifier hyperparameters:

- *n\_estimators*: 100 (number of trees)
- *max\_depth*: 10 (maximum depth of trees)
- *random\_state*: 42 (for reproducibility)
- Validation: 5-fold cross-validation during training

**Performance Evaluation:** Test set metrics (80-20 split):

- Accuracy: 0.731 (73.1% correct predictions)
- Precision: 0.755 (75.5% of positive predictions were correct)
- Recall: 0.667 (66.7% of actual positives were identified)
- $F_1$  Score: 0.708 (harmonic mean of precision and recall)

Analysis: The balanced metrics indicate good overall performance without significant bias. The model shows stronger precision than recall, suggesting conservative predictions.

	Predicted Negative	Predicted Positive
Actual Negative	<i>TN</i> : 3245	<i>FP</i> : 856
Actual Positive	<i>FN</i> : 987	<i>TP</i> : 2912

Table 4.1: Confusion Matrix for Cardiovascular Disease Prediction

#### 4.4.2 Diabetes Disease Prediction

The diabetes disease prediction model is implemented in the DiabetesPredictor class, trained on a dataset of 100,000 patient records.

**Feature Handling:**

- Required features:
  - Age (in years, range: 0-120)



- Gender (categorical: male/female)
- BMI ( $kg/m^2$ , range: 10 – 60)
- Blood glucose level (mg/dL, range: 50-400)
- Optional features:
  - Hypertension (0: absent, 1: present)
  - Heart disease (0: absent, 1: present)
  - Smoking history (never, former, current)
  - HbA1c level (percentage, range: 4.0-15.0)

**Data Preprocessing:**

- Missing value handling:
  - Binary features (hypertension, heart disease) default to 0
  - Categorical features (smoking history) default to 'never'
  - Numerical features (HbA1c) default to population mean
- Feature transformations:
  - Label encoding for gender:  $\{male : 0, female : 1\}$
  - One-hot encoding for smoking history categories
  - BMI validation:  $weight/height^2$  within valid range
- Feature scaling:
  - Standardization using StandardScaler:  $z = (x - \mu)/\sigma$
  - Applied to: age, BMI, glucose level, HbA1c
  - Scaling parameters preserved for inference

**Model Configuration:** Random Forest Classifier hyperparameters:

- *n\_estimators*: 100 (number of trees)
- *max\_depth*: 10 (maximum depth of trees)
- *random\_state*: 42 (for reproducibility)

- Validation: 5-fold cross-validation during training

**Performance Evaluation:** Test set metrics (80-20 split):

- Accuracy: 0.972 (97.2% correct predictions)
- Precision: 1.000 (100% of positive predictions were correct)
- Recall: 0.674 (67.4% of actual positives were identified)
- $F_1$  Score: 0.805 (harmonic mean of precision and recall)

Analysis: The perfect precision (1.000) combined with lower recall (0.674) indicates a highly conservative model that never makes false positive predictions but misses some positive cases. This behavior reflects class imbalance in the training data and suggests potential for recall improvement.

	Predicted Negative	Predicted Positive
Actual Negative	$TN$ : 4521	$FP$ : 0
Actual Positive	$FN$ : 128	$TP$ : 265

Table 4.2: Confusion Matrix for Diabetes Disease Prediction

### 4.4.3 Risk Level Classification

Both models employ identical probability thresholds for risk level categorization:

**Threshold Definitions and Clinical Impact:**

- Low Risk:  $probability < 0.2$ 
  - Standard monitoring recommendations
  - Annual check-up suggestions
- Moderate Risk:  $0.2 \leq probability < 0.5$ 
  - Increased monitoring frequency
  - Lifestyle modification recommendations
  - Quarterly check-up suggestions
- High Risk:  $0.5 \leq probability < 0.8$ 
  - Regular monitoring requirements
  - Specific lifestyle interventions

- Monthly check-up recommendations
- Very High Risk:  $probability \geq 0.8$ 
  - Immediate medical consultation advised
  - Intensive monitoring protocol
  - Weekly check-up recommendations

**Threshold Selection Rationale:** These thresholds were determined through sensitivity analysis and clinical considerations:

- Lower threshold (0.2):
  - Optimized for early detection
  - 95% sensitivity for serious conditions
  - False positive rate: 15% (accepted for safety)
- Middle thresholds (0.5):
  - Balanced sensitivity (85%) and specificity (90%)
  - Clear delineation between risk categories
  - Corresponds to clinical intervention points
- Upper threshold (0.8):
  - 98% specificity for severe cases
  - Immediate intervention trigger point
  - False positive rate  $< 2\%$

The selected thresholds optimize the balance between early intervention and clinical resource utilization. This stratification enables appropriate urgency in health recommendations while minimizing alert fatigue from false positives. Regular threshold validation is performed using new clinical data to ensure continued effectiveness.

#### 4.4.4 Health Improvement Suggestions

When a user's health risk level is identified as Moderate Risk (0.2-0.5), High Risk (0.5-0.8), or Very High Risk ( $> 0.8$ ), the chatbot provides personalized health improvement suggestions. These suggestions aim to guide the user toward reducing their risk level and improving overall well-being.

=====

Both models provide risk assessments categorized into four levels: Low Risk (probability  $< 0.2$ ), Moderate Risk (0.2-0.5), High Risk (0.5-0.8), and Very High Risk ( $> 0.8$ ). These predictions are integrated into the conversation system, allowing RoboBuddy to provide personalized health insights and recommendations based on the user's risk levels. For instance, users with elevated cardiovascular risk might receive suggestions for blood pressure monitoring, while those with diabetes risk factors could get reminders about blood glucose testing.

#### 4.4.5 Health Improvement Suggestions

When a user's health risk level is identified as Moderate Risk (0.2-0.5), High Risk (0.5-0.8), or Very High Risk ( $> 0.8$ ), the chatbot provides personalized health improvement suggestions. These suggestions aim to guide the user toward reducing their risk level and improving overall well-being.

Suggestions are selected based on:

- Health Profile Data: Age, medical history, and lifestyle factors stored in the application.
- Current Risk Factors: Specific results from cardiovascular or diabetes models.

### 4.5 Sentiment Analysis on User's Message

#### 4.5.1 Overview

The sentiment analysis component monitors the user's emotional state during interactions with the chatbot, leveraging the BERT model *bhadresh-savani/bert-base-go-emotion*. This enables real-time classification of messages into several sentiments.

#### 4.5.2 Methodology

- Message Processing: Each user message, typed or transcribed, is analyzed for sentiment using a pretrained BERT model (<https://huggingface.co/bhadresh-savani/bert-base-go-emotion>). It is trained with a number of 169208 examples on 3 epochs, with a batch size of 16. The dataset used for pretraining contains 58,000 carefully curated Reddit comments labeled for 27 emotion categories ([https://huggingface.co/datasets/google-research-datasets/go\\_emotions](https://huggingface.co/datasets/google-research-datasets/go_emotions)).
- Sentiment Logging: Sentiment scores are stored for individual messages, enabling trend tracking over time. A sentiment score is the value resulted from running the model on an input as a conversation. There are 27 categories of sentiments. Some of them are: admiration, amusement,

anger, annoyance, approval, confusion, curiosity, embarrassment, joy, optimism, relief, surprise and many more.

- Sentiment Logging: Sentiment scores are stored for individual messages, enabling trend tracking over time.

### **4.5.3 Daily Sentiment Tracking**

- Daily Summary: Sentiment logs are aggregated daily to record the distribution of messages, along with an average sentiment score.
- User Monitoring: These records help track long-term emotional trends and provide insights into changes in mental well-being.

### **4.5.4 Overall Sentiment Calculation**

- Session Analysis: At the end of each chat session (or end of day), sentiment scores are aggregated to calculate an overall sentiment score, summarizing the user's emotional state. This step includes having all the sentiments for every message already computed and the overall sentiment score is defined as the predominant sentiment from the list of sentiments computed during the day.
- Alerts and Insights: Negative sentiment trends may trigger system notifications or recommendations to caregivers.

## Chapter 5

# Application (Study Case)

### 5.1 App Description and Main Functionalities

GoldenMate is a mobile app for elderly users featuring RoboBuddy, a conversational AI. Key functionalities include:

- Conversational AI that provides companionship and asks meaningful follow-up questions.
- Health monitoring, where users can enter metrics (heart rate, blood sugar) and receive personalized recommendations.
- Games like trivia and Sudoku to keep users mentally stimulated.

### 5.2 App Design

The application consists of various components, such as a conversation interface, health monitoring, and game modules. Use cases and class diagrams are provided in Figures ? and ?.

### 5.3 Implementation

The app is built using Ionic React for the frontend and Python (Flask) for the backend. NLP models from Hugging Face and custom ML algorithms were used for the conversational and recommendation systems.

### 5.4 Testing of the Codebase

TODO

## Chapter 6

# Conclusion and Future Work

GoldenMate demonstrates the potential for AI in elderly care. Our approach, focused on empathetic conversations and health insights, highlights key strengths in accessibility and personalization. Future work will involve expanding the app's games, improving conversational depth, and adding multilingual support.

# Bibliography

[https://celiatecuida.com/en/home\\_en/](https://celiatecuida.com/en/home_en/)