JOGGING ACTIVITY RECOGNITION USING k-NN ALGORITHM.

AFIFAH BINTI ISMAIL

FACULTY OF COMPUTING AND INFORMATICS

UNIVERSITI MALAYSIA SABAH

2022

JOGGING ACTIVITY RECOGNITION USING k-NN ALGORITHM.

AFIFAH BINTI ISMAIL

THESIS SUBMITTED IN PARTIAL FULFILLMENT FOR THE DEGREE BACHELOR OF COMPUTER SCIENCE WITH HONOURS (NETWORK ENGINEERING)

FACULTY OF COMPUTING AND INFORMATICS
UNIVERSITI MALAYSIA SABAH

2022

VERIFICATION

NAME	**:**	AFIFAH BINTI ISMAIL
MATRIC NUMBER	**:**	BI17110127
TITLE	**:**	JOGGING ACTIVITY RECOGNITION USING k-NN ALGORITHM
DEGREE	**:**	COMPUTER SCIENCE WITH HONOR (NETWORK ENGINEERING)
VIVA'S DATE	**:**	25 JANUARY 2022

CERTIFIED BY;

1. **SUPERVISOR** Signature

   SAMRY @ MOHD SHAMRIE SAININ

DECLARATION

I, hereby declare that the material in this project is my own except for quotations, equations, summaries, and references, which have been duly acknowledged.

14th January 2022

Afifah

AFIFAH BINTI ISMAIL

BI17110127

ACKNOWLEDGEMENT

Firstly, thanks to you to Allah SWT for giving me chance and making my work easy in many ways. I am highly indebted to my supervisor Shamrie Bin Sainin for their guidance and constant supervision as well as for providing necessary information regarding the project also for their support in completing the project. I would like to express my gratitude towards my family for their kind co-operation and encouragement which help me in the completion of this project. I would like to express my special gratitude and thanks to industry persons for giving me such attention and time. My thanks and appreciations also go to my colleague in developing the project and people who have willingly helped me out with their abilities.

AFIFAH BINTI ISMAIL

14 JANUARY 2022

LIST OF ABBREVIATIONS

SDLC	Software Development System
k-NN	k Nearest Neighbor
HAR	Human Activity Recognition
IT	Information Technology
GPX	GPS (Global Positioning System) Exchange Format
UMS	University Malaysia Sabah
RC	Red circle
GS	Green star
BS	Blue star
SVM	Support Vector Machine
IoT	Internet of Things
G-sensor	Gyroscope sensor
RAM	Rando-access memory
IDE	Integrated Development Environment
ERD	Entity Relationship Diagram
CSV	Comma Separate Value
SAD	System Analysis and Designs

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Jogging activity recognition using the k-NN algorithm is a system that can help users collect information data of user speed movement using speed sensor and give the classification of jogging activity to the user. The objective of this project are 1) to investigate human activity recognition (HAR) for jogging activity and k-Nearest Neighbors (k-NN) algorithm for jogging classifier, 2) to apply HAR AND k-NN for jogging recognition and classification and, 3) to test the functionality of the k-NN algorithm of jogging recognition and classification. The prototype contains 10 GPX data that will be used as jogging activity and classify the intensity of jogging activity into running, running easy, jogging, and jogging easy. To recognize and classify the level of jogging intensity, k-Nearest Neighbours (k-NN) algorithms will be considered as a machine learning method. The k-NN algorithm is a simple and easy-to-implement supervised machine learning algorithm that can be used to solve both classification and regression problems. This system will give the user classification of the jogging activity after the information data is processed. Whereas the methodology for this project is using Software Development Life Cycles (SDLC). There are five phases Requirement gathering and analysis, System design, Implementation Integration, testing, and lastly is maintenance. Finally, usability testing will be used for evaluation. The jogging recognition technology is incorporated into a web-based system using PHP and Python after extraction, training, and test of the data complete, to create the working implementation that can classify the user's jogging activity. The output from this project is the system is sometimes unable to predict the jogging activity. The finding in this project is the k-NN algorithm is good feature extraction and classifier. However, to approach the limitation in this project, different feature extraction approaches and the study of additional classifiers, as well as research by training the model with a larger dataset and using more different intensities are needed.

ABSTRAK

PENGECEMAN AKTIVITI JOGING MENGGUNAKAN ALGORITMA k-NN

Pengecaman aktiviti jogging menggunakan algoritma k-NN merupakan satu sistem yang dapat membantu pengguna mengumpul data maklumat pergerakan kelajuan pengguna menggunakan sensor kelajuan dan memberi klasifikasi aktiviti jogging kepada pengguna. Objektif projek ini adalah 1) untuk menyiasat pengecaman aktiviti manusia (HAR) untuk aktiviti joging dan algoritma k-Nearest Neighbors (k-NN) untuk pengelas jogging, 2) untuk mengaplikasikan HAR DAN k-NN untuk pengecaman dan pengelasan jogging dan, 3) untuk menguji kefungsian algoritma k-NN pengecaman dan pengelasan joging. Prototaip tersebut mengandungi 10 data GPX yang akan digunakan sebagai aktiviti jogging dan mengklasifikasikan intensiti aktiviti jogging kepada berlari, lari mudah, jogging dan jogging mudah. Untuk mengenali dan mengklasifikasikan tahap keamatan joging, algoritma k-Nearest Neighbors (k-NN) akan dianggap sebagai kaedah pembelajaran mesin. Algoritma k-NN ialah algoritma pembelajaran mesin terselia yang mudah dan mudah dilaksanakan yang boleh digunakan untuk menyelesaikan masalah klasifikasi dan regresi. Sistem ini akan memberikan klasifikasi pengguna bagi aktiviti jogging selepas data maklumat diproses. Manakala metodologi projek ini adalah menggunakan Software Development Life Cycles (SDLC). Terdapat lima fasa seperti pengumpulan dan analisis Keperluan, reka bentuk Sistem, Penyepaduan Pelaksanaan, dan ujian, dan terakhir ialah penyelenggaraan. Akhir sekali, ujian kebolehgunaan akan digunakan untuk penilaian. Teknologi pengecaman jogging dimasukkan ke dalam sistem berasaskan web menggunakan PHP dan Python selepas pengekstrakan, latihan dan ujian data lengkap, untuk mencipta pelaksanaan kerja yang boleh mengklasifikasikan aktiviti joging pengguna. Output daripada projek ini adalah sistem kadangkala tidak dapat meramalkan aktiviti jogging. Dapatan dalam projek ini ialah algoritma k-NN adalah pengekstrakan ciri dan pengelas yang baik. Walau bagaimanapun, untuk mendekati had dalam projek ini, pendekatan pengekstrakan ciri yang berbeza dan kajian pengelas tambahan, serta penyelidikan dengan melatih model dengan set data yang lebih besar dan menggunakan lebih banyak keamatan berbeza diperlukan.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Introduction

This chapter will discuss the problem background jogging activity recognition with recommendation system, problem statement, objectives as well as the scope of this project.

1.2 Problem Background

As Network Engineering students, we have learned so many things about computer components including software and hardware. However, there is still more to learn about this in this course. Everything that has been taught in this course is very useful in real life because today people live in an advanced IT world. Everyone knows that they cannot live without the internet in this progressively digital world. It is all about connecting people. In this course, we learned about networking, database, programming, and so on. All the knowledge from the syllabus can be applied to this Final Year Project which is compulsory for Computer Science students so that students complete the degree. Students already learned about system analysis and design, writing coding in programming, computer network and management, system management, and others. Using the knowledge of computer science, the title that has been decided and developed is "Jogging activity recognition using the k-NN algorithm".

According to the research, many people hate or cannot do a sport. But most people chose this sport as the only sport that they can do which is jogging. Jogging activity does improve aerobic fitness and cardiovascular health. Plus, it burns calories and can build strength. But many people overlook the jogging intensity. Each intensity level affects a runner's fitness differently. For UMS athletes and students who take the Sports Science program, this system can help them recognize and classify the intensity level of every run.

1.3 Problem Statement

In sports, such as jogging, the wearable smartwatch is available which provides various information such as speed (pace), cadence, and strides. However, most of the smartwatches do not automatically recognize activity and provide intervention, suggestions based on what should be the suitable intensity of the run or activity. The person must always look at the watch and try to judge their activity intensity. Therefore, jogging activity recognition using machine learning can be designed to support HAR. The outcome of the project can be a personal trainer for general use.

1.4 Project Objective

The goal of this project is to help user to track their jogging activity and help them to get recommendations from the system according to their level of jogging intensity. The objectives of this project are:

i. To investigate human activity recognition (HAR) for jogging activity and k-Nearest Neighbors (k-NN) algorithm for jogging classifier.

ii. To apply HAR AND k-NN for jogging recognition and classification.

iii. To test the functionality of the k-NN algorithm of jogging recognition and classification.

1.5 Project Scope

i. To collect GPX data from open street maps with running and jogging tags.

ii. This project scope focuses on jogging activity recognition for UMS students of the Faculty of Psychology and Education that take Sport Science Program and UMS athletes from UMS Sports Centre.

iii. Jogging plan scope focusing on 5K beginner or 30 minutes training based on Beginners Running Plan on the wiki.

1.6 Organization of Report

For the full report, it will contain 7 chapters which are: -

i. Chapter 1 (Introduction)

This chapter gives an overview of the project with the problem background, problem statement, objectives, and project scope highlighted.

ii. Chapter 2 (Literature Review)

This chapter about the project introduces the extensive study done on the topic related especially WSN. This chapter also consists of a definition of terms used throughout the report.

iii. Chapter 3 (Methodology)

This chapter highlighted the brief methodology that will be carried out to complete the project. This also shows the required tools and equipment to be used in this project. This chapter also explains the flow of procedures taken through the project.

iv. Chapter 4 (System Analysis and Design)

This chapter is about the system and analysis of the project. This chapter included the design of devices hardware and the application web which used in this project.

v. Chapter 5 (Implementation)

This chapter is about the implementation of the project which is discussing how the device

system is being configured. This chapter also includes how the installation of hardware and software is done.

vi. Chapter 6 (Testing and Analysis)

This chapter is about the testing and analysis of the system prototype. Plus, the result also will be explained in terms of the limitation of the current system and what must be done to the limitation in future work. This chapter also concludes the testing and result.

vii. Chapter 7 (Conclusion)

In this chapter, all the progress of the project will be concluded generally. This chapter also discussed the flow of the project from the implementation phase until the test and analysis of the result.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter describes the literature review that relates to this research. It discusses human activity recognition and k-NN. It also discusses the related work on human activity recognition using the k-NN algorithm.

2.2 Human Activity Recognition

Human activity recognition, (HAR) is a broad field of study concerned with identifying the specific movement or action of a person based on sensor data. Movements are often typical activities performed indoors, such as walking, running, talking, standing, and sitting (J. Brownlee, 2018). Human activity recognition, or HAR, is a challenging time series classification task. It involves predicting the movement of a person based on sensor data and traditionally involves deep domain expertise and methods from signal processing to correctly engineer features from the raw data to fit a machine learning model (J. Brownlee, 2018). K-NN algorithm as machine learning has achieved accuracy in predicting human activity from the raw sensor data using the classification method. To achieve the accuracy of prediction, we need to train and test the dataset. This project only focused on jogging and running as a human activities. The raw GPX data was collected from an open street website as a dataset in this project.

2.2.1 Jogging Activity Recognition

The system will focus on jogging activity as HAR. When people do a sports activity, two aspects will be their focus which is health and fitness. For health purpose is to not get injured while doing a sport and may release stress while doing it. For fitness purposes, the sports activity also can be challenging enough depending on the intensity and distances. Jogging activity is not too dangerous and fit enough for people to do it. The raw GPX data with running and jogging tags will be collected as jogging activity recognition from an open street website as a dataset in this project. The training data will be created as training data based on the data that will be extracted from the raw GPX data.

2.3 k-NN Algorithm as Machine Learning

K-nearest neighbor is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). KNN has been used in statistical estimation and pattern recognition since the beginning of the 1970s as a non-parametric technique. (Saed Sayad, 2021). For this system, k-NN will be used as the classification of jogging activity. The reason for using this algorithm as a machine learning for this project is because easy to use, has quick calculation time, and the data are not assumed (Madison Schott, 2019).

2.4 k-NN Method

KNN method by finding the distances between a query and all the examples in the data, selecting the specified number examples (K) closest to the query, then votes for the most frequent label (in the case of classification) or averages the labels (in the case of regression). (0. Harrison, 2018)

2.4.1 Measuring Functions

Finding the distances between the query and all examples in the data can be calculated by using Euclidean, Minkowski, or Manhattan. This measuring function is the most popular measuring function to be used in the k-NN algorithm. Distance measuring functions can be taken in the following ways. We can define distance function d (x, y) E between two points by measuring their distance according to the Euclidean formula or d (x, y). (S. Kaghyan, 2012). Manhattan distance is calculated as the sum of the absolute differences between the two vectors. Minkowski distance calculates the distance between two real-valued vectors. The distance function that measures the distance between them using formulas below:

Distance functions

Euclidean $\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$

Manhattan $\sum_{i=1}^{k}|x_i - y_i|$

Minkowski $\left(\sum_{i=1}^{k}(|x_i - y_i|)^q\right)^{1/q}$

2.4.2 The optimal k-Values

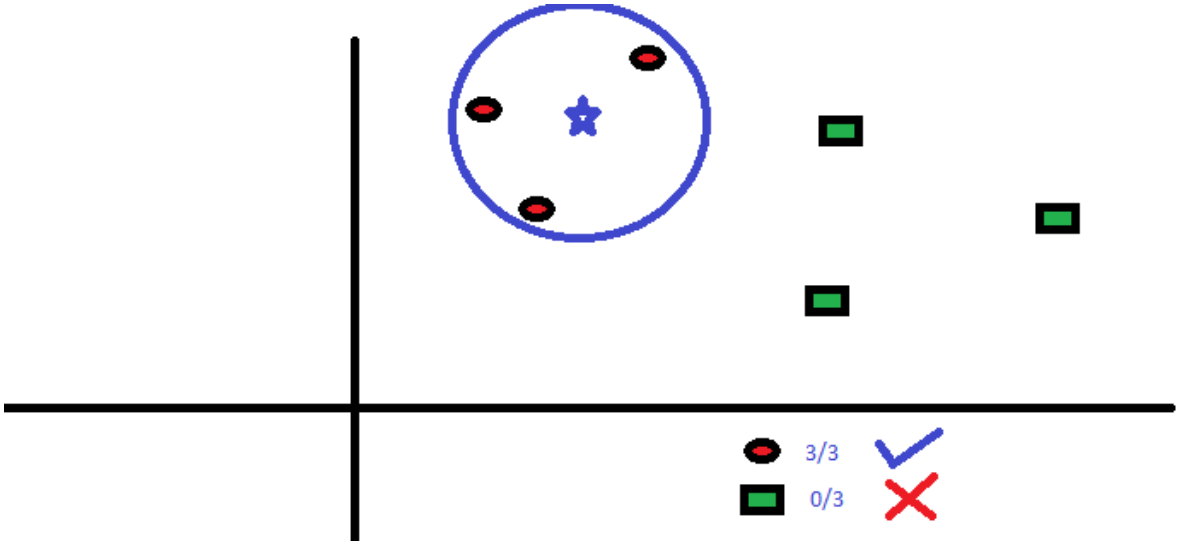

Figure 2.1 Figuring k- Values

The following is the spread of red circles (RC) and green squares (GS). The purpose is to find the class blue star (BS) either in RC or GS. The value K is the nearest neighbor that similar thing exists near each other. For example, K=3 means three data points that are near BS and BS is a center. The tree points all RC, so BS is a class of RC. (T. Srivastava, 2108)

2.4.3 Classification

The K-nearest neighbors (K-NN) for classification, uses a similar idea to the KNN regression. For KNN, a unit will be classified as most of its neighbors. The KNN algorithm for classification will look at the k nearest neighbors of the input then the output is the most frequent label among those k examples. (A. Teixeira- Pinto, 2018)

2.4.4 Regression

K-NN regression is a non-parametric method that, intuitively, approximates the association between independent variables and the continuous outcome by averaging the observations in the same *neighborhood*. The size of the neighborhood needs to be set by the analyst or can be chosen using cross-validation (we will see this later) to select the size that minimizes the mean-squared error. (A. Teixeira- Pinto, 2018)

2.5 k-NN on Jogging Activity Recognition

K-Nearest Neighbor is a supervised learning algorithm where the result of a new instance query is classified based on the majority of the K-Nearest Neighbor category. (S. Kahyan, 2012) Nonetheless, human physical activity recognition has been used as the input feature. In this study, we consider jogging activities as HAR. The reason for selecting these activities was quite simple. This activity was selected because jogging is one of the sports activities is performed regularly by many people. K-NN will be used as a classifier after the data of jogging activity has been processed.

2.6 Related Work: Human Activity Recognition using k-NN Algorithm

i. Human Activity Analysis using Machine Learning Classification Techniques
This study aims to evaluate the efficacy of different machine learning classification algorithms Smartphones, both low-cost and industrial, are used as sensors to monitor human activities. The aim is to test machine learning algorithms on real-world datasets to assess their accuracy and draw useful conclusions. A comparative study was performed among the applied various techniques k-NN, SVM, Random Forest, Neural Networks, Logistic regression, and Naïve Bayes. In them, Logistic Regression and the neural network gave good results whereas Naive Bayes's result was not good. The limitation of this work is though the efficiency of the neural network is good, the model is not dynamic. (Z. Gulzar et.al, 2019)

ii. IoT-Based Activity Recognition with Machine Learning from Smartwatch

This project proposed that the location of the arm, because of arm movement, may not be as precise as other areas of the body. To recognize the activity, they used a smartwatch as a new IoT comfortable wearable. The researcher investigated a smartwatch as a platform with accuracy and F1-score parameters and they compared algorithms by python. For recognition, they considered k-Nearest Neighbors (k-NN) and Decision Tree (DT) (N. Mozaffari et.al, 2020). The results show using a smartwatch as a new solution can be accurate 99%. Using the k-NN algorithm for classification, an accurate activity recognition system that utilizes a triaxial accelerometer on the waist has shown 95% to 98% of accuracy (3) and on the chest with 97.9% accuracy (4). The limitation, this paper only recognize activity such as walking and jogging.

iii. Smartphone sensor-based activity recognition by using machine learning and deep learning algorithms.

The study aims to create an experiment for recognizing human activity using data from smartphone sensors. Based on the accelerometer and gyroscope sensor data, they implemented a dataset and developed two types of features to identify human activity and smartphone motion. The research shows that the proposed movements are successful in detecting the target. Binary classification of the events was performed to produce better and more effective results when compared to performing individual events as it results in a negligible amount of inaccuracy When the detection accuracy was evaluated, the combination of events resulted in a higher error rate. Based on the findings, it can be concluded that accelerometer sensor (A-sensor) readings contribute more than G-sensor readings (Gyroscope sensor), but that using both can boost detection accuracy. Machine learning and deep learning algorithms were used to evaluate their performance. The findings also showed that the results were consistent across a range of mobile device orientations. Both internal sensors are used to measure the phone's movements. (Q. Liu et. al, 2018)

2.7 Available Jogging Application

A closely related application project is the STRAVA app that is available in the Google store apps store and website. This app is designed for users to track their running activity. There are several features available in this app such as tracking users' swimming and cycling activity. Also, the apps can manage weekly goals.

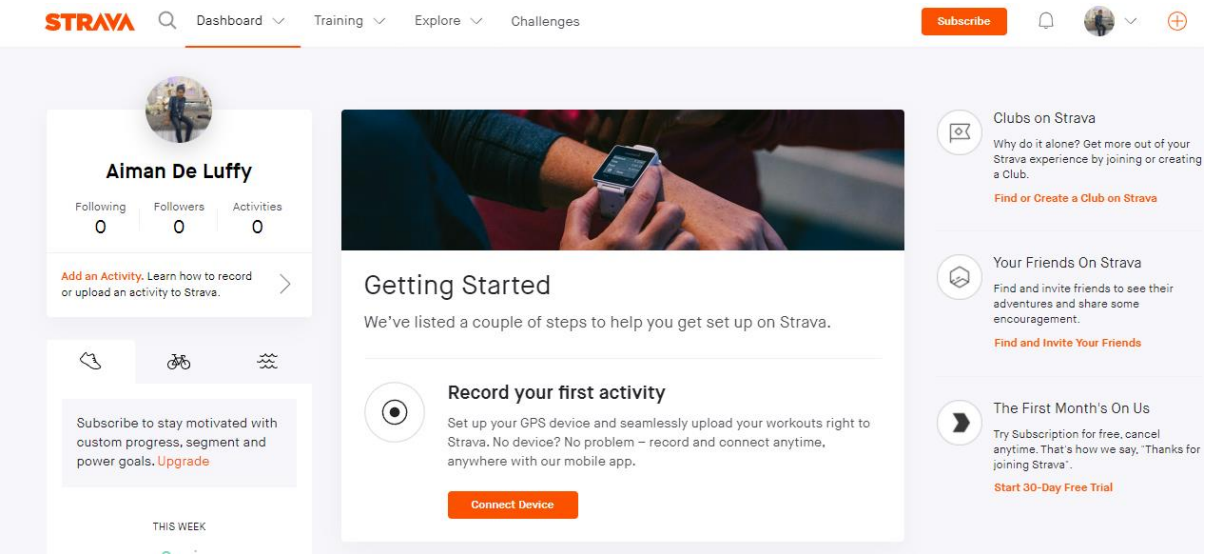

Figure 2.2 Strava Website

2.8 Conclusion

Based on the review in chapter 2, human activity recognition is to identify the movement of humans such as walking, running, and talking. The k-NN algorithm is a machine learning to classify and predict the accuracy of the dataset. Then, a review of between the existing systems to be a guideline in developing this system.