MARMARA
UNIVERSITY

MARMARA UNIVERSITY

FACULTY OF ENGINEERING

CSE 4062 - INTRODUCTION TO DATA SCIENCE AND ANALYTICS

2021-2022 SPRING TERM

INSTRUCTOR: Doç.Dr. Murat Can GANİZ

PREDICTING THE MOST EXPERIENCED AND EXPEDITIOUS WORKERS ON THE REQUESTED ISSUES

GROUP 1:

Beyza İlayda TEMUÇİN	**150619842**	**Chemical Engineering**	**ilaydatemucin@marun.edu.tr**
Beyza AYDOĞAN	**150117030**	**Computer Engineering**	**beyzaaydogan99@gmail.com**
Güneş YÜZAK	**150117004**	**Computer Engineering**	**gunesyuzak@gmail.com**
Merve MÖREY	**150617046**	**Chemical Engineering**	**mervemorey@hotmail.com**
Mücahit Mustafa USTA	**150416020**	**Mechanical Engineering**	**m.mustafa_5091@hotmail.com**

Delivery Date: 26.04.2022

PROJECT DESCRIPTION

IT services are significant for almost all companies, therefore the companies providing this service should develop a suitable system for solving the requested issues. Appointment of the right employee for the right job has a great contribution in making everything more solution-oriented and faster. The main point aimed to achieve in our project is to determine the most experienced and expeditious workers on the requested issues and to accelerate the wheels of the company.

Our project consists of two prediction tasks with three selection steps. The first step is to predict the appropriate worker for the requested issue among the 14 class labels. The number of solved issues according to the given class label will be found in order to select the most experienced 3 workers from 102 workers on requested issue. Then in the second step, it is necessary to obtain each of them average work logs for the requested issue. The third step is to predict the worker with the least average time spent on the requested issue and this worker's average time. Consequently, we focus on both experience and return speediness.

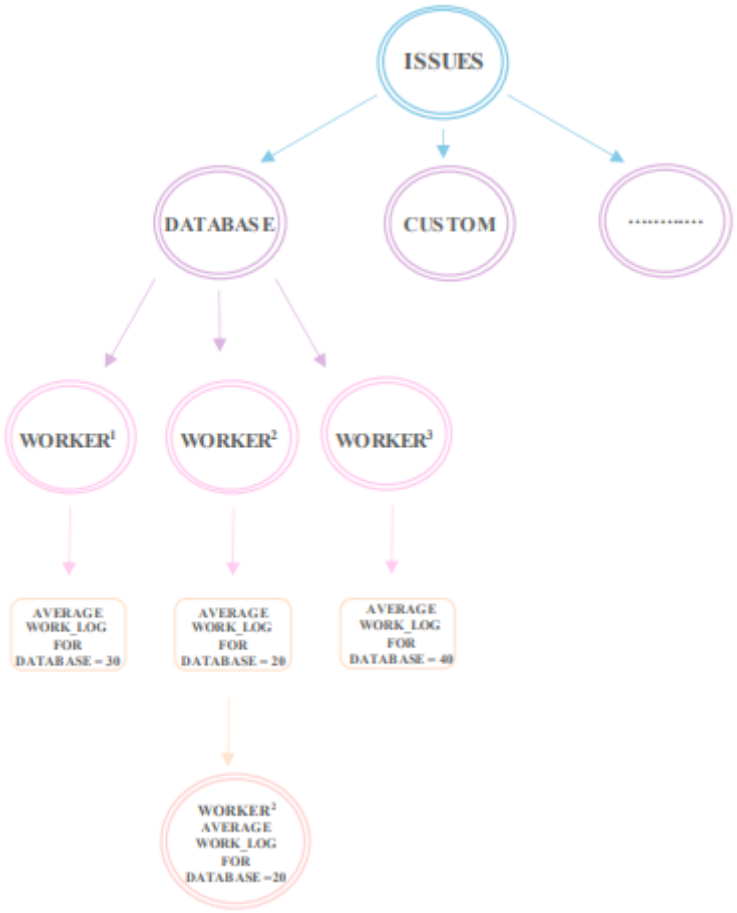

Figure 1: Process Flow Chart

DATA STATISTICS

In our dataset “Issue Tickets”
Numbers of columns → 11
Number of rows → 16970

In the beginning we had a data table with 16 columns, which are Issue_Id, Reporter, Issue_Type, Priority, Compname, Jiraname, Worker, Employee_Type, Work_Log, Work_Log_Total, Work_Log_Ratio, Issue_Category, Issue_Sub_Category, Label, Creation_Date and Resolution_Date. We deleted Issue_Id, Work_Log_Total, Work_Log_Ratio, Issue_Sub_Category, Label features to progress in a specific way.
The important features are Worker, Work_Log, Issue_Category.
The type of Worker is nominal, Work_Log is numeric and Issue_Category is nominal.
We chose Issue_Category as class attribute, Work_Log for regression and Worker for classification as target attribute.