



PES UNIVERSITY, BANGALORE

Department of Computer Science and Engineering

Attrition Analysis System

By
K S Ramalakshmi
Sri Charan Kanumuri
Kruthi S
Kshitij G Saha



PES UNIVERSITY, BANGALORE

Department of Computer Science and Engineering

Table of Contents

Introduction	4
Purpose	4
Intended Audience	4
Product Scope	4
References	4
Overall Description	4
Product Perspective	4
Product Functions	4
User Classes and Characteristics	5
Operating Environment	5
Design and Implementation Constraints	5
Assumptions and Dependencies	5
External Interface Requirements	6
User Interfaces	6
Software Interfaces	6
Communications Interfaces	6
Analysis Models	7
System Features	10
Attrition Prediction	10
Exploratory Data Analysis	10
Other Nonfunctional Requirements	11
Performance Requirements	11
Safety Requirements	11
Security Requirements	11
Software Quality Attributes	11
Business Rules	11
Other Requirements	12
Project Planning Document	12
Lifecycle	12
Tools to be used	12
Deliverables	12
Work Breakdown Structure	12
Gantt Chart	13
Testing	13
User Stories	14
Code	14



PES UNIVERSITY, BANGALORE

Department of Computer Science and Engineering

Revision History

Name	Date	Reason For Changes	Version
Attrition Analysis Software	31/10/22	Implemented Flask Backend	1.1.0



PES UNIVERSITY, BANGALORE

Department of Computer Science and Engineering

Introduction

Purpose

The product for which the software requirements are specified in this document is the Attrition Analysis System, version 1.0.

The document elaborates on the

- Need for the software
- Software and other requirements
- Design and Implementation Constraints
- Features aimed to implement in version 1.0

AAS 1.0 will be deployed as a web-app for the end user, with the main features implemented.

Intended Audience

The document is intended for the product developers, project manager, testers, documentation writers and team mentor.

Product Scope

The Attrition analysis system helps the Product and Service Industries to analyze patterns and potential trends that affect attrition rates in companies. A thorough analysis on these trends help companies understand underlying reasons behind attrition and can take measures to mitigate the situation and retain valuable employees. Being an effective HR analytics tool, the software helps businesses in making informed decisions with respect to managing its employees and ensuring qualitative outcome.

References

Dataset for training is used from Kaggle.com. The application is suited to run in any recent browser hosting a server.

Overall Description

Product Perspective

The product aims to provide business insights into hiring employees and ensure their continued service during their tenure. The various trends predicted navigate the organization's investments and efforts into ensuring the employee's smooth tenure.

Product Functions

Main features the products aims to provide:



PES UNIVERSITY, BANGALORE

Department of Computer Science and Engineering

- Attrition prediction
- Employee trend analysis in the organization

User Classes and Characteristics

- Employers
- Product Managers
- Board Members of an organization

Operating Environment

The software operates with a server (preferably Live/Apache Xampp). No particular hardware requirements. The app can run in any operating system provided it is run on a latest updated browser.

Design and Implementation Constraints

The software performs analysis and prediction only on a fixed set of attributes.

2.6 Assumptions and Dependencies

The backend for analysis assumes the same set of attributes to be tested on employees against attributes tailored for each organization.

External Interface Requirements

User Interfaces

The web application comes with a simple HTML front end with a provision to upload a CSV file. The UI is fairly minimalistic for the user with a single file upload. After processing of the file, the results along with statistics are displayed on the web page itself.

Software Interfaces

The application is hosted using a Flask server which communicates with a TensorFlow backend using implicit Python module import. Version 1 accepts CSV format of input only. Python packages used for the server implementation and the backend include Numpy, Pandas, Matplotlib, Sklearn, Tensorflow, Flask, Seaborn, werkzeug, and OS. Frontend is implemented using the basic web stack containing HTML, CSS and JavaScript.

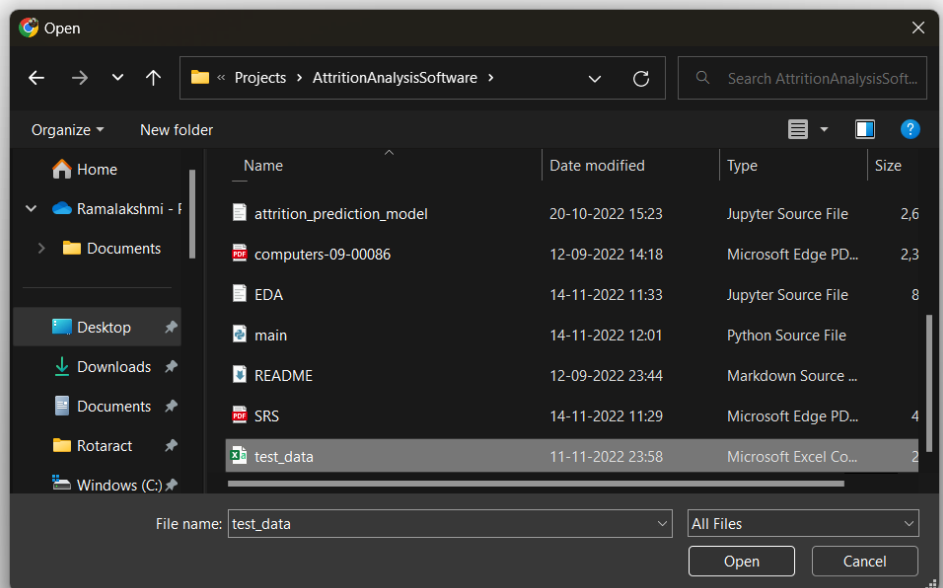


PES UNIVERSITY, BANGALORE

Department of Computer Science and Engineering

Attrition Prediction Software

No file chosen



Analysis Models

Artificial Neural Networks are utilized to analyze and further predict the likelihood of attrition in companies. Based on the attributes provided by the customer including but not limited to age, sex, department, Education field etc. The model includes one input layer, one hidden layer and one output layer. The input, hidden and output layer contain 6,6,1 units respectively. The categorical data is encoded to ensure input meets the standards of the Artificial Neural Network. The input and hidden layers have the Rectifier activation function and the output layer has the sigmoid activation function which given a binary output.



PES UNIVERSITY, BANGALORE

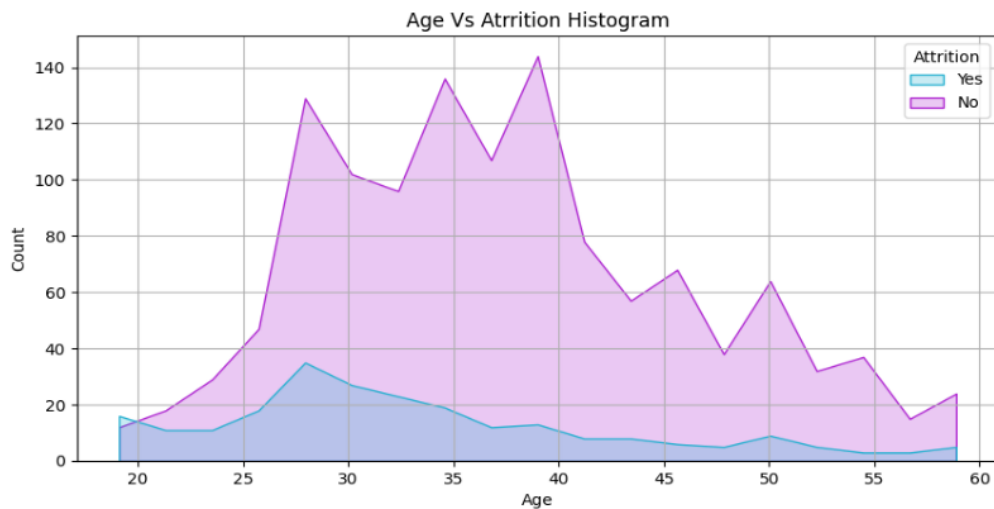
Department of Computer Science and Engineering

Attrition Prediction Results

[EDA Graphs](#)

FromHome	Education	EducationField	EmployeeCount	EmployeeNumber	EnvironmentSat
	2	Life Sciences	1	1	2
	1	Life Sciences	1	2	3
	2	Other	1	4	4
	4	Life Sciences	1	5	4
	1	Medical	1	7	1
	2	Life Sciences	1	8	4
	3	Medical	1	10	3
	1	Life Sciences	1	11	4
	3	Life Sciences	1	12	4

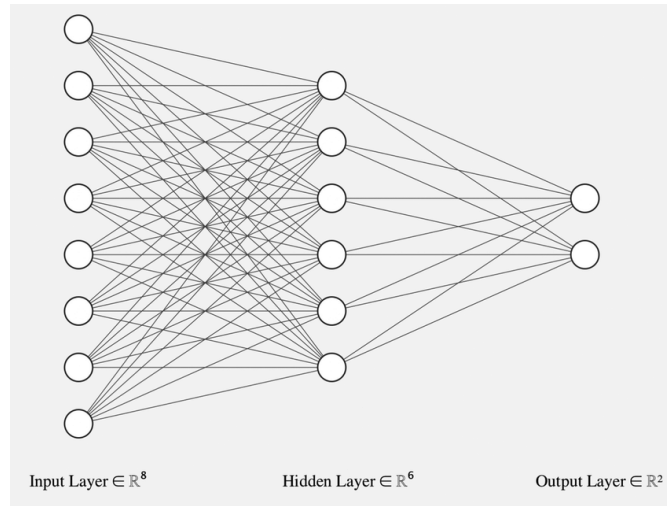
Miscellaneous Graphs





PES UNIVERSITY, BANGALORE

Department of Computer Science and Engineering



Sample ANN

System Features

Attrition Prediction

5.1.1 Description and Priority

Priority: High

The software predicts whether a given set of employees will go through attrition or not, depending on a trend analysis from a previous employee dataset.

5.1.2 Stimulus/Response Sequences

The user needs to upload a csv file, no matter the size, and for the given file, we predict 'yes' or 'no' w.r.t attrition for each employee.

5.1.3 Functional Requirements

REQ-1: Latest Chrome Browser to support Live Server

REQ-2: Updated Python 3.10.4 and dependent libraries

Exploratory Data Analysis

5.2.1 Description and Priority

Priority: High

The software performs an exploratory data analysis on the input dataset based on pre-defined metrics and outputs the trends identified from the same as EDA plots.

5.2.2 Stimulus/Response Sequences



PES UNIVERSITY, BANGALORE

Department of Computer Science and Engineering

The user needs to upload a csv file, no matter the size, and for the given file, we perform exploratory data analysis for the whole batch of employees.

5.1.3 Functional Requirements

REQ-1: Latest Chrome Browser to support Live Server

REQ-2: Updated Python 3.10.4 and dependent libraries

Other Nonfunctional Requirements

Safety Requirements

The website will consist of RSA 2048 encryption to ensure secure communication and secure access of data to prevent data tampering.

Security Requirements

Confidential data shared by the customer will include explicit details about their employees. This data will be extracted by the web application and will be given as input for Exploratory Data Analysis and into the Artificial Neural Network. Customers must agree to the privacy policy giving access to the confidential employee database. The database and the information it entails will not be shared with any third party under any circumstances. Additional layers of security will ensure that the data is not accessed by eavesdroppers.

Software Quality Attributes

The software is based on javascript web application and python libraries including matplotlib and tensorflow, keras. These are easily adaptable and maintainable. In addition, availability and accessibility is taken care of since the product is a web application available on the internet. The web-app is system independent and interoperable.

Other Requirements

Privacy policy to agree to the sharing of employee databases

Project Planning Document

Lifecycle

The lifecycle chosen for the execution of the project is Scrum. It is our understanding that with a legacy lifecycle, the plan would be rigid and firm, useful for very large project as they are easy to enforce. But Scrum allows us to be little more flexible in terms of our planning and execution, especially when it is a small team executing this project.

Tools to be used

We will be using JIRA primarily to track our progress and planning. The code will be hosted and managed using Git and GitHub. Naturally, we will be using them for version control of the project. Coding will be done mainly on VSCode for the website portion and Google Colab



PES UNIVERSITY, BANGALORE

Department of Computer Science and Engineering

for EDA and machine learning model construction. MySQL or MongoDB will be the databases of choice to store data as far as the website is concerned. Figma will be used to create a wireframe for the website. Bug tracking will be done using JIRA again with a combination of a spreadsheet.

Deliverables

1. EDA on the uploaded dataset (include graphs) - Once the user uploads a dataset with a fixed schema, there will be a certain set of standard operations and the results of these operations will be displayed on the user's screen for them to infer.
2. Prediction on the uploaded dataset – After the uploading of the dataset, the user can look at the predictions made by the model, based on which the user may take further action as he/she deems necessary.

Work Breakdown Structure (WBS)

The work will divide broadly in 3 Sprints which are,

1. Sprint 1 – Build necessary components
 - a. Write a script for a standard EDA
 - b. Build a ML model for prediction
 - c. Design wireframe for the website
2. Sprint 2 – Build a basic website
 - a. Setup backend of the website with a database
 - b. Construct frontend of the website
 - c. Integrate EDA script and model in the website
3. Sprint 3 – Complete the Website
 - a. Improve the frontend of the website where ever possible
 - b. Host website on AWS along with database and model

Each sprint will last for an average of 3 weeks.

Gantt Chart

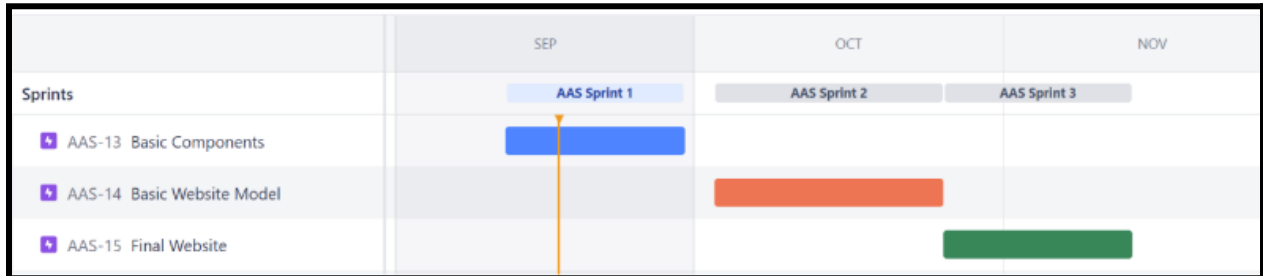
The following is a rough Gantt Chart of how things are expected to go:

- a. Basic Components – 12-Sep to 29-Sep
- b. Basic Website Model – 3-Oct to 25-Oct
- c. Final Website – 26-Oct to 13-Nov

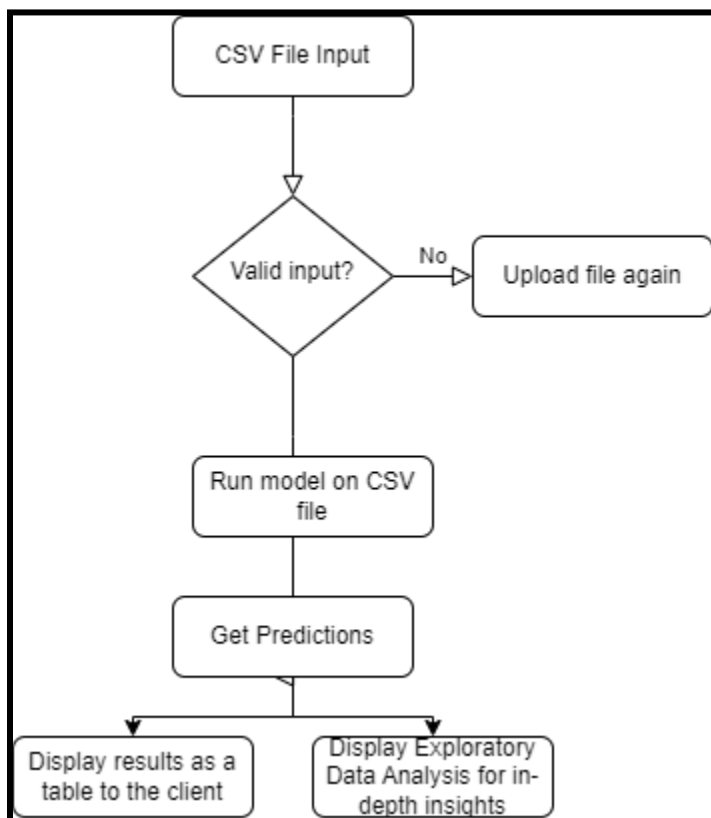


PES UNIVERSITY, BANGALORE

Department of Computer Science and Engineering



Architecture Design of the project and Data Flow Diagram





PES UNIVERSITY, BANGALORE

Department of Computer Science and Engineering

User Stories

Managing a Team

“As a project manager, I want an attrition prediction software, to select the right targets to achieve work goals, and know the current mindset and working goals of my team. This is quintessential for leading an organization, to set future goals and define the heading path. Attrition helps me get a rough idea of the current state of my employees to continue in the team, measure their dedication and perseverance to the work and help define a smaller subgroup to define goals and execute plans to achieve the same.”

Select the Right Team

“As a Recruiter, I want an attrition prediction software in the initial stages of recruitment to know the long-standing reliability to ensure a resilient team is formed, with a very less dropout rate since expenditure and associated expenses accompany an employee’s tenure at the company.”

Code

All codes, modules and other resources used in the software were regularly updated, committed and pushed to a GitHub repository from the start. The link to the same has been mentioned below:

<https://github.com/ksramalakshmi/AttritionAnalysisSoftware.git>

Testing

The software was tested with varied inputs and below are the following results obtained

Test 1

An input CSV file with 10,000 records containing the same number and type of columns as the train CSV data.

Result: Test case passed with 91% accuracy and corresponding EDA graphs were also obtained



PES UNIVERSITY, BANGALORE

Department of Computer Science and Engineering

Test 2

An input CSV file with 10,000 records containing different numbers and types of columns from the train CSV data.

Result: Test case failed, model and EDA predictions did not run.