



FINAL YEAR PROJECT REPORT

STARTUPFATE: PREDICTING SUCCESS FOR STARTUPS WITH MACHINE LEARNING

**In fulfillment of the requirement
For degree of
BS (COMPUTER SCIENCES)**

By

**HASSAN RASOOL
MUHAMMAD HARIS
RAYYAN AHMED**

**70109 (BSCS)
70161 (BSCS)
70164 (BSCS)**

SUPERVISED

BY

MR. MUHAMMAD SHAHID KHAN

BAHRIA UNIVERSITY (KARACHI CAMPUS)

SPRING-2024

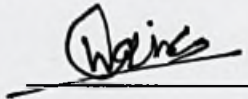
DECLARATION

We hereby declare that this project report is based on our original work except for citations and quotations which have been duly acknowledged. We also declare that it has not been previously and concurrently submitted for any other degree or award at Bahria University or other institutions.

Signature :  _____

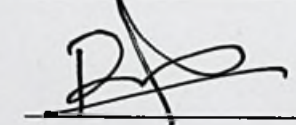
Name : Hassan Rasool

Reg No. : 70109

Signature :  _____

Name : Muhammad Haris

Reg No. : 70161

Signature :  _____

Name : Rayyan Ahmed

Reg No. : 70164

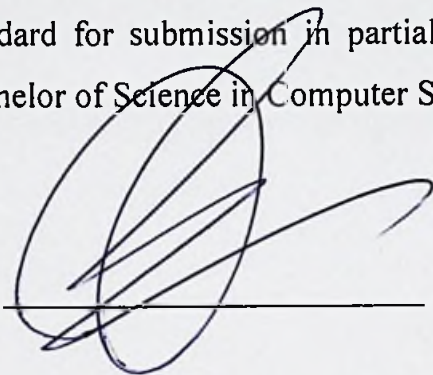
Date : 12-6-2024

APPROVAL FOR SUBMISSION

I certify that this project report entitled **“STARTUPFATE: PREDICTING SUCCESS FOR STARTUPS WITH MACHINE LEARNING”** was prepared by **HASSAN RASOOL, MUHAMMAD HARIS and RAYYAN AHMED** has met the required standard for submission in partial fulfilment of the requirements for the award of Bachelor of Science in Computer Science at Bahria University, Karachi.

Approved by,

Signature :



Supervisor: Mr. Muhammad Shahid Khan

Date :

12/06/24

The copyright of this report belongs to Bahria University according to the Intellectual Property Policy of Bahria University BUORIC-P15 amended on April 2019. Due acknowledgement shall always be made of the use of any material contained in, or derived from, this report.

© 2024 Bahria University. All right reserved.

STARTUPFATE: PREDICTING SUCCESS FOR STARTUPS WITH MACHINE LEARNING

ABSTRACT

Startup failure is a significant concern in the business landscape, with most new ventures facing closure within their first few years. Identifying factors that contribute to startup failure can assist investors in making informed decisions, entrepreneurs in mitigating risks, and policymakers in creating supportive environments. This project aims to address this problem by leveraging historical startup data and developing a predictive model to identify startups at risk of failure. We have shown that Artificial Intelligence can leverage publicly available data on the Internet to calculate the probability of each of these outcomes with a high level of confidence.

The idea for this project came to us considering the failure of Pakistan's largest startup, "Airlift." Witnessing the rise and fall of such a significant player in the startup landscape prompted us to introspect, aiming to not only learn from their missteps but also to contribute something unique and valuable to the entrepreneurial ecosystem. Our project aims to develop an Artificial Intelligence model capable of predicting the likelihood of a startup's success or failure based on relevant features and factors. This project culminates in the creation of a user-friendly website where users can input startup information and receive predictions, aiding entrepreneurs, and investors in making informed decisions.

TABLE OF CONTENTS

DECLARATION	ii
APPROVAL FOR SUBMISSION	iii
ABSTRACT	v
TABLE OF CONTENTS	vi
LIST OF TABLES	x
LIST OF FIGURES	xi
LIST OF SYMBOLS / ABBREVIATIONS	xiii
LIST OF APPENDICES	xiv

CHAPTERS

1	INTRODUCTION	1
	1.1 Background	1
	1.2 Problem Statement	1
	1.3 Aims and Objectives	2
	1.4 Scope of Project	2
2	LITERATURE REVIEW	4
	2.1 Background	4
	2.2 Introduction to Startups	4
	2.3 Importance of Predicting Startup Success	5
	2.4 Start-Up Definition & Growing Importance	6
	2.5 Key Features for Startup Success Prediction	7
	2.6 Machine Learning	8
	2.7 Machine Learning in Startup Success Prediction	9
	2.8 Deep Learning	10
	2.9 Neural Networks for Enhanced Prediction	11
	2.10 Why We Use Neural Networks for Startup Success Prediction	12
	2.11 Advantages of ANN for Startup Success Prediction	12
	2.12 Application of ANN in StartupFate	14
	2.13 Related Work	14

2.14	Comparison Table with Existing Study	16
2.15	Chapter Summary	17
3	REQUIREMENT SPECIFICATIONS	19
3.1	Existing Systems	19
3.1.1	Description	19
3.1.2	Limitations	19
3.2	Proposed System	20
3.2.1	Overview	20
3.2.2	Features	20
3.2.3	Advantages	21
3.3	Requirement Specifications:	21
3.3.1	Functional Requirements	21
3.3.2	Non-Functional Requirements	22
3.4	Use Cases	22
3.4.1	Use Case 1: Predict Startup Success	22
3.4.2	Use Case 2: View Resources	23
3.5	Chapter Summary	23
4	DESIGN AND METHODOLOGY	24
4.1	Data Acquisition and Pre-Processing	24
4.1.1	Data Acquisition	24
4.1.2	Data Pre-Processing	24
4.2	Methodology (Framework/Architecture)	25
4.3	Data Sources	26
4.4	Dataset Creation	26
4.4.1	Cleaning	26
4.4.2	Preprocessing	26
4.5	Exploratory Data Analysis (EDA)	27
4.6	Model Development	33
4.6.1	Training Methodology	33
4.6.2	Model Training	33
4.7	Model Comparison	34

4.7.1	Machine Learning Models	34
4.7.2	Gradient Boosting (XG-Boost)	35
4.7.3	Random Forest Classifier	35
4.7.4	Multinomial Logistic Regression	36
4.7.5	Decision Tree	36
4.7.6	Support Vector Machines (SVM)	37
4.7.7	Logistic Regression	37
4.8	Deep Learning Models	38
4.8.1	Evaluation Metrics	39
4.8.2	Confusion Matrix	39
4.8.3	Accuracy	40
4.8.4	Precision	40
4.8.5	Recall	40
4.8.6	F1-Score	40
4.8.7	Classification Performance	41
4.9	Process Model	42
4.10	Gantt Chart	43
4.11	Sequence Diagram	43
4.12	Web Interface	44
4.13	Chapter Summary	50
5	SYSTEM IMPLEMENTATION	51
5.1	Frontend Implementation	51
5.1.1	Prediction Panel	51
5.1.2	Resources Panel	52
5.2	Backend Implementation	52
5.2.1	Model Implementation	52
5.3	Workflow	53
5.4	Workflow Diagram	54
5.5	Web Flow Diagram	55
5.5	Chapter Summary	55

6	SYSTEM TESTING AND EVALUATION	56
6.1	Introduction	56
6.2	Testing Process Overview	56
6.3	Significance of Testing in Software Development	56
6.4	Test Planning	56
6.5	Test Cases	57
6.6	Testing Approaches and Techniques	61
6.6.1	Automated Testing	61
6.6.2	Usability Testing	62
6.6.3	Unity Testing	62
6.6.4	Integration Testing	62
6.6.5	End-to-End Testing	62
6.6.6	User Acceptance Testing (UAT) i	62
6.7	Evaluation Metrics	62
6.7.1	Accuracy	62
6.7.2	Performance	63
6.7.3	Reliability	63
6.7.4	User Satisfaction	63
6.8	Chapter Summary	63
7	CONCLUSION	64
7.1	Summary of Work	64
7.2	Significance and Impact	64
7.3	Expand Training Data	64
7.4	Deployment and Business Validation	65
7.5	Model Interpretability	65
7.6	Continual Learning	65
7.7	Chapter Summary	65
	REFERENCES	66