



SCHOOL OF COMPUTER SCIENCE AND APPLICATIONS

A Project Report

On

Air Pollution Forecasting using LSTM and GRU

Submitted in partial fulfillment of the requirements for the award of the Degree of

Master of Computer Science

Submitted by

Davada Juned Aslam

R20SCS06

under the guidance of

Internal Guide

Dr. Hemant K. S.

July 2022

Rukmini Knowledge Park, Kattigenahalli, Yelahanka, Bengaluru-560064

www.reva.edu.in



SCHOOL OF COMPUTER SCIENCE AND APPLICATIONS

CERTIFICATE

The project work titled - **Air Pollution Forecasting using LSTM and GRU**, is being carried out under our guidance by **Davada Juned Aslam, R20SCS06** , a bonafide student of REVA University, and is submitting the project report in partial fulfillment, for the award of **MSc Computer Science** during the academic year **2021–22**. The project report has been approved, as it satisfies the academic requirements with respect to the Project Work prescribed for the aforementioned Degree.

Signature with Date

Hemant K. S.
Internal Guide

Signature with Date

Prof. Vijayalakshmi .K
Academic Vertical Head-PG

Signature with Date

Dr. S. Senthil
Professor and Director

Name of the Examiner with Affiliation

Signature with Date

- 1.
- 2.

DECLARATION

I, Mr. Davada Juned Aslam, pursuing my **MSc of Computer Science**, offered by School of Computer Science and Applications, REVA University, declare that this Project titled - “Air Pollution Forecasting using LSTM and GRU” , is the result of the Project Work done by me under the supervision of Dr. Hemanth K. S. (Associate Professor).

I am submitting this Project Work in partial fulfillment of the requirements for the award of the degree of Master of Computer Applications by REVA University, Bengaluru, during the Academic Year 2021-22.

I further declare that this Project Report or any part of it has not been submitted for the award of any other Degree / Diploma of this University or any other University/ Institution.

(Signature of the candidate)

Signed by me on:

Certified that this project work submitted by Davada Juned Aslam has been carried out under our guidance and the declaration made by the candidate is true to the best of my knowledge.

Signature of Internal Guide

Date :

Signature of Director of the School

Date :

Official Seal of the School

ACKNOWLEDGEMENT

I hereby acknowledge all those, under whose support and encouragement, I have been able to fulfil all my academic commitments successfully. In this regard, I take this opportunity to express my deep sense of gratitude and sincere thanks to School of Computer Science and Applications which has always been a tremendous source of guidance.

I express my sincere gratitude to **Dr. P. SHYAMA RAJU**, Honorable Chancellor, REVA University, and Bengaluru for providing us the state-of-the-art facilities.

I am thankful to **Dr. M. DHANAMJAYA**, Vice Chancellor and **Dr. N. RAMESH**, Registrar, REVA University, Bengaluru for their support and encouragement.

I take this opportunity to express my heartfelt thanks to **Dr. S. SENTHIL**, Professor & Director, School of CSA, REVA University and my sincere thanks to **Prof. VIJAYALAKSHMI.K** Associate Professor, Academic Vertical Head for PG Programs, School of CSA, REVA University, whose encouragement and best wishes provided impetus for the Project Work carried out.

Also, my sincere gratitude goes to my **Dr. Hemant K. S. Associate Professor**, for the valuable suggestion and constant encouragement towards the completion of this project.

Last, but not the least, I thank my parents for their incredible support and encouragement throughout.

ABSTRACT

With the development of the industry in the last few years. We are facing an issue related to air quality. We are not getting the proper air quality because of the pollution out there in the environment due to harmful gases from the industry. It will affect human health and it can cause a serious issues related to the lungs. Air pollution can cause by different ways in today's world like the CO₂ released from the car and some harmful chemicals which are released in the air by the industry. Air pollution can spread with the flow of air means wind direction and speed. To overcome this real-time problem there is much research going on nowadays that can forecast air pollution but it required a lot of computational power. Our approach in this research is that we are going to forecast air pollution using the very famous deep learning technique Recurrent Neural Networks (RNN) based framework with special structure memory cell known as Long Short-term memory (LSTM) and Gated Recurrent Unit (GRU). We can easily forecast air pollution using this easily by just providing some last day's data to the model. It will forecast the next 24 hours of data by just providing the last 15 days of air pollution of data.

TABLE OF CONTENTS

Sr. No.	CHAPTERS	PAGE NO
1.	INTRODUCTION	1
	1.1 INTRODUCTION TO PROJECT	
	- SOFTWARE AND HARDWARE SPECIFICATION	
2.	LITERATURE SURVEY	4
3.	SYSTEM ANALYSIS	7
	3.1 EXISTING SYSTEM	
	3.2 LIMITATIONS OF THE EXISTING SYSTEM	
	3.3 PROPOSED SYSTEM	
	3.4 ADVANTAGES OF THE PROPOSED SYSTEM	
4.	DATASET	8
5.	SYSTEM DEVELOPMENT STRATEGY	17
	4.1 METHODOLOGY	
	4.2 IMPLEMENTATION	
6.	SYSTEM DESIGN	36
	5.1 HIGH LEVEL DESIGN (ARCHITECTURAL)	
7.	RESULT	37
8.	CONCLUSION AND SCOPE FOR FUTURE ENCHANCEMENT	42
9.	BIBILOGRAPHY	43

IMAGE INDEX

IMG No.	CAPTION	PAGE NO
1.1	Represent the number of death cause by different diseases in the world	2
4.1	Data Visualization for Pollution column	9
4.2	Data Visualization for Dew column	10
4.3	Data Visualization for temp column	11
4.4	Data Visualization for Press column	12
4.5	Data Visualization for Wnd_spd column	13
4.6	Data Visualization for Snow column	14
4.7	Data Visualization for Rain column	15
4.8	Sample data for the dataset	16
5.1.1	The LSTM architecture	18
5.1.2	The Forget Gate for the LSTM	20
5.1.3	Showing the range of sigmoid activation function (0-1)	20
5.1.4	The Input gate of the LSTM	21
5.1.5	Comparing the range of activation function and sigmoid activation function	22
5.1.6	Output Gate for the LSTM architecture	22
5.1.7	The GRU architecture	23
5.1.8	Update gate for the GRU	24
5.1.9	Forget gate for the GRU	25
5.2.1	Represent of data variable from dataset	26
5.2.2	Dataset before normalizing data	28

5.2.3	Data visualization for the Air Pollution on the hourly basis for the last 15 days which is 360 hours of data	29
5.2.4	Heat map for the co-relation of data of the dataset	30
5.2.5	The dataset visualization after normalizing the data	31
5.2.6	The Normalized data for the prediction variable pollution	31
7.1	The graph for the 300 epoch values for the GRU with 50 neurons	37
7.2	The forecasting of data for the next 24 hours data based on the data what we have provided for 50 neurons	38
7.3	The graph of 100 and 200 epoch values for the 50 neurons for the GRU	38
7.4	The graph for the both the model with 100 and 200 epoch values.	39
7.5	The train and test validation data for the model with 50 neurons and 300 epochs	39
7.6	The output graph for the next 24 hours of data for the 50 neurons	40
7.7	The training graph for the 100 and 200 epochs with 50 neurons	40
7.8	The output prediction graph for the actual and predicted values	41

TABLE INDEX

TAB No.	CAPTION	PAGE NO
4.1	Range of data and count for the Pollution Column	8
4.2	Min, Max, Average, SD for the Pollution Column	8
4.3	Range of data and count for the Dew Column	9
4.4	Min, Max, Average, SD for the Dew Column	9
4.5	Range of data and count for the Temp Column	10
4.6	Min, Max, Average, SD for the Temp Column	10
4.7	Range of data and count for the Press Column	11
4.8	Min, Max, Average, SD for the Press Column	11
4.9	Range of data and count for the Wind direction Column	12
4.10	Range of data and count for the Wnd_spd Column	13
4.11	Min, Max, Average, SD for the Wnd_spd Column	13
4.12	Range of data and count for the Snow Column	14
4.13	Min, Max, Average, SD for the Snow Column	14
4.14	Range of data and count for the Rain Column	15
4.15	Min, Max, Average, SD for the Rain Column	15
5.2.1	Model Architecture for GRU	32
5.2.2	Model Architecture for LSTM	33