

# AI Research Navigation - A Scholarly Journey

Shantakumar B. Patil

# AI Research Navigation - A Scholarly Journey

**Shantakumar B. Patil**

Sai Vidya Institute of Technology, Rajanukunte, Karnataka, India



**DeepScience**

*Published, marketed, and distributed by:*

Deep Science Publishing, 2025  
USA | UK | India | Turkey  
Reg. No. MH-33-0523625  
[www.deepscienceresearch.com](http://www.deepscienceresearch.com)  
[editor@deepscienceresearch.com](mailto:editor@deepscienceresearch.com)  
WhatsApp: +91 7977171947

ISBN: 978-93-7185-574-7

E-ISBN: 978-93-7185-286-9

<https://doi.org/10.70593/978-93-7185-286-9>

Copyright © Shantakumar B. Patil, 2025.

**Citation:** Patil, S. B. (2025). *Artificial Intelligence Research Navigation - A Scholarly Journey*. Deep Science Publishing. <https://doi.org/10.70593/978-93-7185-286-9>

This book is published online under a fully open access program and is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0). This open access license allows third parties to copy and redistribute the material in any medium or format, provided that proper attribution is given to the author(s) and the published source. The publishers, authors, and editors are not responsible for errors or omissions, or for any consequences arising from the application of the information presented in this book, and make no warranty, express or implied, regarding the content of this publication. Although the publisher, authors, and editors have made every effort to ensure that the content is not misleading or false, they do not represent or warrant that the information-particularly regarding verification by third parties-has been verified. The publisher is neutral with regard to jurisdictional claims in published maps and institutional affiliations. The authors and publishers have made every effort to contact all copyright holders of the material reproduced in this publication and apologize to anyone we may have been unable to reach. If any copyright material has not been acknowledged, please write to us so we can correct it in a future reprint.

## Preface

The expanse of AI has flourished into a foundation for contemporary research which influences many fields in computer science. However, the scope of AI has led to enormous number of challenges for foundation theory of any innovation. The AI Research Navigation; A Scholarly journey book sets one's sight on imparting systematic and organized perspective on AI related research. The book is intended for providing academicians and researchers not just a glimpse of AI but a detailed description of case studies in the relevant field for further research expansion. The intention of authors is not just navigating through different perspectives of Artificial intelligence but to cultivate an interest in readers the potential to interrogate, associate and contribute sensibly to unravel the story of Artificial intelligence

Dr. Shantakumar B. Patil

# Table of Contents

---

Chapter no.	Chapter Name	Page no.
<b>1</b>	<b>DIVE INTO DATA COLLECTION</b>	<b>1</b>
	<b>1.1 Introduction</b>	<b>1</b>
	1.1.1 Methods of data collection	2
	1.1.2 Secondary data gathering guide	4
	1.1.3 Secondary data types	4
	1.1.4 Secondary data formats	5
	1.1.5 Secondary data analysis methods	6
	1.1.6 Datasets	9
	1.1.7 Understanding the importance of custom datasets	10
	1.1.8 Ethical considerations in dataset creation	10
	<b>1.2 Dataset aggregator and Providers</b>	<b>11</b>
	1.2.1 Sources of datasets	11
	1.2.2 Open Datasets for Natural Language Processing	13
<b>2</b>	<b>DATA CONSOLIDATION AND REPORTING</b>	<b>16</b>
	<b>2.1 Introduction</b>	<b>16</b>
	2.1.1 Data Preprocessing: An Overview	16
	2.1.2 Data Quality: Why Preprocess the Data?	16
	2.1.3 What is Data Aggregation?	17
	2.1.4 Importance and Benefits of Data Aggregation	17
	2.1.5 Data Preprocessing for Aggregation	18
	<b>2.2 Data Cleaning</b>	<b>18</b>
	2.2.1 Missing Values	18
	2.2.2 Noisy Data	20
	2.2.3 Data Cleaning as a Process	22
	<b>2.3 Data Aggregation Techniques</b>	<b>25</b>
	2.3.1 Summarization Techniques	25
	2.3.2 Grouping and Categorization	25
	2.3.3 Time-Based Aggregation	25
	2.3.4 Data Aggregation in Different Domains	26
	2.3.5 Data Aggregation Examples	27

	<b>2.4 Roles of Machine Learning Models</b>	<b>28</b>
	2.4.1 Machine Learning Regression Models	29
	2.4.1.1 Regression Metrics	29
	2.4.1.2 Examples of Regression Evaluation Metrics	30
<b>3</b>	<b>CONFUSION MATRIX: A GATEWAY TO AI MODEL EVALUATION</b>	<b>37</b>
	3.1 Introduction to confusion matrix	37
	3.2 Confusion Matrix for Binary Classes	38
	3.3 Confusion Matrix for Multiple Classes	41
	3.4 Tools for Computing a Confusion Matrix	48
<b>4</b>	<b>EXPLORING CLASSIFICATION TECHNIQUES IN AI</b>	<b>53</b>
	4.1 Introduction to AI classification techniques	53
	4.2 Logistic Regression	53
	4.2.1 Significance of Logistic regression	53
	4.3 Naive Bayes	57
	4.4 K-nearest Neighbors algorithm	60
	4.4.1 Introduction	60
	4.4.2 Working of KNN algorithm	60
	4.4.3 Decision Tree	63
	4.4.4 Role of Decision Trees	64
	4.5 Random Forest	66
	4.5.1 Random Forest Classifier	66
	4.6 Support Vector Machines	68
	4.6.1 Advantages and Disadvantages of SVM	69
	4.6.2 Types of SVM	70
	4.7 Artificial Neural Networks	71
	4.7.1 Benefits of ANN	71
	4.7.2 Types of ANN	72
	<b>Annexure</b>	<b>75</b>
	<b>Case Study of Research</b>	
<b>5</b>	AI-Systems To Strengthen, Enhance Customer Trust And Mitigate Risks In And Digital Transactions	<b>80</b>
<b>6</b>	The Language Of Machines: LLMs As Catalysts In Academic AI Exploration	<b>85</b>
<b>7</b>	Medical Decision Support System For The Missing Data Using Data Mining	<b>91</b>

## List of Figures

<b>Figure no.</b>	<b>Figure Name</b>	<b>Page no.</b>
1.1	Sources of data collection	1
2.1	2-D plot of customer data	21
3.1	Confusion Matrix	37
3.2	Number of values evaluated	40
3.3.1	Confusion matrix obtained by training a classifier	42
3.3.2	Converting the matrix to a one-vs-all matrix	43
3.3.3	One-vs-all matrix	43
3.3.4	Receiver Operating Characteristics curve for different case	46
3.3.5	ROC curve for a binary classification problem	47
3.3.6	ROC curve for various dataset	48
3.4.1	Binary confusion matrix	49
3.4.2	Confusion matrices for the multi-class data	50
3.4.3	FPR, TPR and the threshold values	51
4.2.1	Advantages of Logistic Regression	54
4.3.1	Probability table for weather condition and playing sports	59
4.4.1	Visualization of KNN algorithm	60
4.4.2.1	Dataset of two classes	61
4.4.2.2	Representation of new data entry to dataset	62
4.4.2.3	Representation of last data entry	62
4.4.3.1	Sample decision tree for sports example	63
4.4.3.2	Splitting of decision nodes	65
4.5.1.1	Basic diagram of Random forest classifier	67
4.5.1.2	Simple decision tree for price of phone	67
4.6.1	Hyperplane separates the tags	68
4.6.2	Best Hyperplane	69
4.6.3	SVM classifier example	70
4.7.1	ANN architecture	71
4.7.2.1	Img based nural classifier	73
4.7.2.2	Forecasting	73

## List of Tables

<b>Table no.</b>	<b>Table Name</b>	<b>Page no.</b>
2.4.1.1	Simple house details	29
2.4.1.2	House details with actual rent value	31
2.4.1.3	Difference between the actual and predicted value	31
2.4.1.4	Example of actual demand and forecasted demand for a brand of ice creams	34
2.4.1.5	Comparison of MSE/RSME, MAE and $R^2$	36
3.2.1	Data classified for both cancer and non-cancer individuals	39
3.2.2	Predicted values for both cancer and non-cancer individuals	39
3.3.1	Test set of samples	41
3.3.2	Calculated class-wise accuracy, precision, recall, and f1-score	44
3.3.3	Comparison of TPR and FPR	46
4.2.2.1	Dataset of Brightness and Saturation	63



# **Chapter -1**

## **Dive into Data Collection**

### **Authors**

**Dr. Shantakumar B. Patil**  
Professor,  
Sai Vidya Institute of Technology, Bengaluru, Karnataka, India

**Prof. Manjusha P K**  
Assistant Professor,  
Sai Vidya Institute of Technology, Bengaluru, Karnataka, India

**Dr. Varun E**  
Associate Professor,  
Sai Vidya Institute of Technology, Bengaluru, Karnataka, India

**Prof. H S Poornima Gowda**  
Assistant Professor,  
Sai Vidya Institute of Technology, Bengaluru, Karnataka, India

**Prof. Bhagya M**  
Assistant Professor,  
Cambridge Institute of Technology North Campus,  
Bengaluru, Karnataka, India

**Prof. Prabha Seetaram Naik**  
Assistant Professor,  
Cambridge Institute of Technology North Campus,  
Bengaluru, Karnataka, India