

CONTENTS

Chapter	Contents	Page No
1. INTRODUCTION		1
1.1 Background		1
1.1.1 Temporal data prediction		1
1.1.2 The Problem of Predicting the Future		1
1.2 Motivation for the Research		2
1.3 Definition of the Problem		4
1.4 Objectives and Scope of research work		5
1.5 Methodology		6
1.5.1 Review of Current Research		6
1.5.2 Data Sources		6
1.5.3 Identify and analyze the benefits of predicting techniques		7
1.5.4 Hybrid Model		8
1.5.5 Research on the rising temporal big data prediction		8
1.6 Chapter Organization		9
1.7 Summary		10
2. LITERATURE REVIEW		11
2.1 Introduction		11
2.2 Temporal Data Mining Tasks		11
2.3 Temporal Big data Prediction		12
2.4 The Problem and Potential of Temporal big data prediction		13
2.5 Challenges for Forecasting with Temporal Big Data		14
2.6 Summary		16
3. TEMPORAL BIG DATA MINING		18
3.1 Introduction		18
3.2 Temporal big data		19
3.3 Overview of Temporal big database		20
3.4 Time series		21
3.5 Event Sequences & Episodes		23
3.6 Event-Interval Sequence		24
3.7 Data Streams		26
3.8 Temporal Networks		27

3.9 Spatio temporal Data	28
3.10 Temporal big data mining	28
3.10.1 Representation of temporal data	29
3.10.2 Time-domain based representations	29
3.10.3 Transformation based representations	30
3.10.4 Generative model based representations	30
3.11 Similarity measures	32
3.11.1 Similarity in time	32
3.11.2 Similarity in shape	33
3.11.3 Similarity in change	33
3.12 Mining tasks	34
3.13 Time series Temporal Data Mining	36
3.14 Data Mining Analysis Theory	37
3.15 Summary	38
4. BIG DATA IN DATA MINING FOR FORECASTING	39
4.1 Introduction	39
4.2 Overview, Value Intention and Basics	39
4.3 Defining the Need	40
4.4 Big Data in Data Mining for Forecasting	41
4.5 Selecting and Shrinking Big Data	43
4.6 Temporal data vs. Transactional Modeling	43
4.7 Use of External Data	44
4.8 Process and Methods for Data Mining for Forecasting	45
4.9 Temporal data Based Variable Reduction and Selection	48
4.10 Example for Manufacturing Data Mining Forecasting	50
4.11 Summary	55
5. CLOUD-BASED PLATFORM	56
5.1 The Cloud as an Enabler for Big Data Analytics	56
5.2 Cloud Technologies	56
5.3 Cloud and Big Data: A Compelling Combination	57
5.4 Data Storage	58
5.4.1 Unstructured Data: Object-Based Storage Systems	58
5.4.2 Structured Data: Distributed File Systems	58
5.4.3 Structured Data: Key-Value Stores	59

6.4.2 Model Topology	88
6.4.3 Hybrid Model Implementation	88
6.4.4 Model initialization	89
6.4.5 Hidden Markov Model Result	90
6.5 Training and Prediction	90
6.6 Baum Welch Algorithm	91
6.7 Forward procedure	94
6.8 Backward procedure	95
6.9 Likelihood Based Prediction Algorithm	97
6.10 Prediction	97
6.11 Hybrid Model Algorithm	98
6.12 Summary	98
7. EXPERIMENTATION	99
7.1 Software Requirements	99
7.2 Hardware Requirements	99
7.3 Cloud Environment	100
7.4 Results	101
7.5 Temporal Sock Data Prediction	102
7.6 Testing	104
7.6.1 Scaling Test	105
7.6.2 Data Split Handling Comparison	106
7.7 Summary	107
8. EVALUATION AND RESULTS	108
8.1 Critical Analysis	108
8.2 Training Time	109
8.3 Prediction time	110
8.4 Accuracy	112
8.5 Error rate	113
8.6 Memory Consumption	115
8.7 Hybrid Model Performance	116
8.8 Summary	
9. CONCLUSIONS AND SCOPE FOR FUTURE WORK	118
9.1 Future Work	119
REFERENCES	
ANNEXURE	

LIST OF FIGURES

Figure	Particulars	Page No.
Figure 3.1	Illustrations for some of the mentioned temporal big data	18
Figure 3.2	Time series, Time sequence and Temporal data visualization	20
Figure 3.3	Time cube	21
Figure 3.4	An example of a time series.	22
Figure 3.5	Example of an event-sequence with length.	23
Figure 3.6	Example of an arrangement of length three and size three	25
Figure 3.7	An example where mapping event-interval sequences to discrete event sequences may cause ambiguities	26
Figure 4.1	Examples of externally available temporal data	42
Figure 4.2	The Process for Developing Temporal data Forecasting Models with exogenous variables	47
Figure 4.3	Business Hierarchy	50
Figure 4.4	Operating Segments	51
Figure 4.5	Daisy Chain Approach	52
Figure 4.6	Graphical form	54
Figure 4.7	Visual rendition of Model Driver Contributions	54
Figure 5.1	HDFS Architecture	59
Figure 5.2	BlobSeer Architecture	59
Figure 5.3	AzureMapReduce Architecture	60
Figure 5.4	Amazon EC2 Architecture	64
Figure 5.5	Architecture of Microsoft Azure	65
Figure 6.1	Overview of Hybrid Model	73
Figure 6.2	General flow of MapReduce Technique	77
Figure 6.3	Illustrating the MapReduce Technique	78
Figure 6.4	Hybrid model for prediction of rising temporal data	81
Figure 6.5	Improved K-means clustering for temporal data	82
Figure.6.6	K means and Improved K-means clustering	83
Figure.6.7	Hidden Markov Model	87
Figure 6.8	Hidden Markov Model Initialization Parameters	90
Figure 6.9	Hidden Markov Model Results	91
Figure.6.10	Baum Weltch Model	92

Figure.6.11 Forward Procedure	94
Figure 7.1 Cloud Environment	100
Figure 7.2 Actual and Predicted Stock Price – ICICI Bank Ltd	102
Figure 7.3 Actual and Predicted Stock Price – HCL Technologies	103
Figure 7.4 Actual and Predicted Stock Price – State Bank of India	103
Figure 7.5 Actual and Predicted Stock Price – Larsen & Tourbo Limited	103
Figure 7.6 Actual and Predicted Stock Price – TVS Motors Limited	104
Figure 7.7 Actual and Predicted Stock Price – Maruthi Motors	104
Figure 8.1 Training time of the hybrid model	109
Figure 8.2 Mean training time	110
Figure 8.3 Prediction time	111
Figure 8.4 Mean Prediction time	112
Figure 8.5 Evaluated performance of the system in terms of accuracy prediction	113
Figure 8.6 Error Rate Percentage	114
Figure 8.7 Mean Error Rate Percentage	115
Figure 8.8 Memory Consumption	116

LIST OF TABLES

Table	Particulars	Page No.
Table 4.1	Model approach comparison results	52
Table 4.2	Example of Model quality across the GBU by Area level of the Hierarchy	54
Table 5.1	Cloud Features Comparison	69
Table 6.1	Sample Data Set	75
Table 6.2	Map Reduced Sample Data	79
Table 6.3	Improved K means clustering for temporal data	82
Table 6.4	Cluster-1 (Same)	83
Table 6.5	Cluster-2 (Low)	84
Table 6.6	Cluster-3 (High)	84
Table 6.7	Calculation of transition probabilities	85
Table 6.8	Example for transition probabilities	85
Table 6.9	Observation Matrix	86
Table 6.10	Hybrid Model algorithm	98
Table 7.1	List of Framework Components and their recommended hardware configurations	100
Table 7.2	Stock prices	101
Table 7.3	Test Data	102
Table 7.4	Predicted prices	104
Table 7.5	Azure execution times for scaling test	105
Table 7.6	AZURE Cost estimation	105
Table 7.7	AZURE Price	106
Table 7.8	Computational efficiency of the proposed architecture	106
Table 8.1	MAPE for different stocks	109
Table 8.2	Training time of the hybrid model	109
Table 8.3	Comparative Time	111
Table 8.4	Evaluated performance of the system in terms of accuracy Prediction	113
Table 8.5	Error Rate Percentage	114
Table 8.6	Memory Consumption	115
Table 8.7	Performance Summary	116