

Table of Contents

About the Author	ix
About the Technical Reviewer	xi
Acknowledgements	xiii
Introduction	xv
Chapter 1: An Introduction to Streaming Data	1
Streaming Data	1
The Need to Process and Analyze Streaming Data.....	3
The Challenges of Streaming Data.....	4
Applications of Streaming Data	5
Windowing Techniques	7
Incremental Learning and Online Learning.....	11
Introduction to the Scikit-Multiflow Framework.....	18
Streaming Data Generators.....	19
Create a Data Stream from a CSV file	24
Summary.....	27
References.....	27
Chapter 2: Concept Drift Detection in Data Streams.....	31
Concept Drift.....	31
Adaptive Windowing Method for Concept Drift Detection.....	33
Drift Detection Method.....	37

TABLE OF CONTENTS

- Early Drift Detection Method..... 39
- Drift Detection Using HDDM_A and HDDM_W..... 39
- Drift Detection Using the Page-Hinkley Method..... 50
- Summary..... 53
- References..... 53

- Chapter 3: Supervised Learning for Streaming Data57**
- Evaluation Methods..... 58
- Decision Trees for Streaming Data..... 59
 - Hoeffding Tree Classifier..... 60
 - Hoeffding Adaptive Tree Classifier 63
 - Extremely Fast Decision Tree Classifier..... 67
 - Hoeffding Tree Regressor 70
 - Hoeffding Adaptive Tree Regressor 72
- Lazy Learning Methods for Streaming Data..... 74
- Ensemble Learning for Streaming Data 82
 - Adaptive Random Forests..... 82
 - Online Bagging 86
 - Online Boosting 90
- Data Stream Preprocessing 92
- Summary..... 93
- References..... 94

- Chapter 4: Unsupervised Learning and Other Tools for Data Stream Mining97**
- Unsupervised Learning for Streaming Data 97
 - Clustering 98
 - Anomaly Detection 101

TABLE OF CONTENTS

Other Tools and Technologies for Data Stream Mining	104
Massive Online Analysis (MOA).....	105
Apache Spark	105
Apache Flink.....	106
Apache Storm.....	106
Apache Kafka	107
Faust.....	107
Creme	108
River	108
Conclusion and the Path Forward	111
References.....	112
Index.....	115