**CANDIDATE DECLERATION**

I, Nanak Chand, student of Computer Science and Engineering Department, National Institute of Technical Teachers Training and Research, Chandigarh, bearing Roll Number-142407 submit this pre-thesis work entitled **“Big *Data Analytics using Multi-classifier for Large Datasets*”** submitted to Panjab University, Chandigarh, India declaring the work done is genuine and produced under the guidance of Dr. C. Rama Krishna and Dr. Emmanuel S. Pilli. The contributions of others involved, every effort is honored with due reference to the literature and acknowledgement of discussions fulfilling the intellectual property rights.

I declare that if any violation to the intellectual property rights or copyrights is found then I will be solely responsible for it. My supervisor should not be held for full or partial violation of copyrights.

**Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Place: NITTTR, Chandigarh Nanak Chand**

**CERTIFICATE**

This is to certify that, this pre-thesis report titled **“*Big Data Analytics using Multi-classifier for Large Datasets*”** submitted by Nanak Chand bearing Reg. No.14-TTC-186 is bonafide record of the research work carried out by her in partial fulfillment for the requirement of the award ofMaster of Engineering (Computer Science and Engineering) degree from Panjab University, Chandigarh at National Institute of Technical Teachers Training and Research, Chandigarh. This pre-thesis has not been submitted to any other University or Institution for the award of any degree.

SUPERVISORS

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Dr. C. Rama Krishna Dr. Emmanuel S. Pilli**

**(Internal Guide) (External Guide)**

Professor and Head Assistant Professor

Department of CSE Department of CSE

NITTTR, Chandigarh MNIT, Jaipur

**ACKNOWLEDGEMENT**

All the work produced in here by me is all because that almighty God and my parents bestow their blessing hand on me at every moment and have provided me with the ability, courage, resources, opportunity and their kind support. I would be very thankful to our Director, **Prof. (Dr.) M. P. Poonia,** National Institute of Technical Teachers Training and Research, Chandigarh, for providing us amiable environment and facilities. I am highly indebted to my supervisor **Dr. C. Rama Krishna,** Professor and Head, Department of Computer Science and Engineering, National Institute of Technical Teachers Training and Research, Chandigarh, for his guidance and constant supervision for giving insight into problem and also for his support to complete my project.

I wish to express my deep sense of gratitude to **Dr. Emmanuel S. Pilli**, Assistant Professor, MNIT Jaipur for his stimulating suggestions, for his regular inspiration, scholarly guidance and helpful suggestions throughout the course of this pre-thesis work. I am indebted to him for allowing me to draw upon his precious and valuable time.

I also thanks to all other faculty members and staff of Department of Computer Science, NITTTR Chandigarh for their intellectual support throughout the course of stay at NITTTR, Chandigarh. I would like to express my gratitude to my parents for their unconditional support and treated my pre-thesis priority important.

Nanak Chand

ME (CSE)

Roll Number: 142407

Reg. No: 14-TTC-186

**LIST OF FIGURES**

| **Figure No.** | **Name of the figure** | **Page No.** |
| --- | --- | --- |
| 1.1 | 3 Vs of Big Data | 1 |
| 2.1 | Overview of Logic of Multi-Classifier Systems | 7 |
| 2.2 | Taxonomy of Ensuring Diversification | 9 |
| 2.3 | Classifiers in Cascading Fashion | 10 |
| 2.4 | Parallel Multi-Classifier Topology | 10 |
| 2.5 | Hybrid Topology for Multi-Classifier Systems | 11 |
| 2.6 | Feature-based classification by input values | 12 |
| 2.7 | Feature-based multi-classifier by both input and output of individual classifier | 12 |
| 2.8 | Big Data Processing Framework Proposed By Wu | 15 |
| 2.9 | Big Data Analytics Architecture by Smith | 16 |
| 4.1 | Map Reduce Algorithm for Large Datasets | 21 |
| 4.2 | Hadoop Ecosystem | 22 |

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
| **Table No.** | **Name of the Table** | **Page No.** |

**LIST OF ABBREVIATIONS**

|  |  |
| --- | --- |
| **Abbreviation** | **Description** |
| SVM | Support Vector Machine |
| KNN | K Nearest Neighbor |
| ROC | Receiver Operating Characteristics |
| Auc | Area Under Curve |