**TITLE OF PROJECT**

|  |
| --- |
| **Distributed Subspace Clustering with SPARK** |

**PROJECT MEMBERS** (*Maximum 4 members per group*)

|  |  |
| --- | --- |
| **ENROLLMENT NO/EMAIL** | **PHOTO** |
| AM.EN.U4CSE15055  SURENDHAR SATYAVARAPU  chinnu.satyavarapu@gmail.com | C:\Users\Hp\Desktop\text books gate\1.jpg |
| AM.EN.U4CSE15044  P MUNINDRA NAIDU  nagamunindra2@gmail.com |  |
| AM.EN.U4CSE15144  NITHISH YADAV P  nitheeshkumar122@gmail.com |  |
| AM.EN.U4CSE15017  B SUNIL KUMAR REDDY  nitheeshkumar122@gmail.com |  |

**GUIDE**

|  |
| --- |
| NAME : Ms. Sandhya Harikumar  DEPARTMENT : Computer Science and Engineering  EMAIL : sandhyaharikumar@am.amrita.edu  CONTACT NO : + 91 96335 49831 |

|  |
| --- |
| ***Signed by Guide (with date)*** |

**EXTENSION OF MINOR PROJECT**

YES NO

**AREA OF PROJECT**

|  |
| --- |
| Machine Learning |

**AIM OF THE PROJECT**

|  |
| --- |
| Using Distributed Machine Learning to improve the Subspace clustering algorithm used within several systems. |

**SOFTWARE AND HARDWARE REQUIREMENTS OF PROJECT**

|  |
| --- |
| 1. *Ubuntu* 2. *Apache Spark* 3. *Jupyter Notebook* |

**Synopsis of the Project**

|  |
| --- |
| Usually, machine learning models are applied on the data within the system. But this approach is sluggish when the data is spread across different systems. In this case if we merge all the data into a single system to carry out the algorithms, it is cost intensive and time-consuming process.  Distributed machine learning is helpful in this context. Here we maintain all the machines using a master system which can perform all the machine learning tasks parallely. As the data is already distributed across many systems using this approach can improve the performance of the algorithm and it can be cost effective.  In our project we use distributed machine learning for the subspace clustering algorithms. |

**References for the Project (***Can be the references of papers if doing research, or any websites or materials that you refer***)**

|  |
| --- |
| * **Aggarwal, Charu C., et al. "Fast algorithms for projected clustering." *ACM SIGMoD Record*. Vol. 28. No. 2. ACM, 1999.** * **Meng, Xiangrui, et al. "Mllib: Machine learning in apache spark." *The Journal of Machine Learning Research* 17.1 (2016): 1235-1241.** * **Bekkerman, Ron, Mikhail Bilenko, and John Langford, eds. *Scaling up machine learning: Parallel and distributed approaches*. Cambridge University Press, 2011.** |

***We hereby assure that the project that we will undertake will be done by us to its originality.***

*Signed by Project Members*

*Date:*

**APPROVAL STATUS**

Approved Not approved

**Suggestions for project approval / reformulating the Project:**

***Name and Signature of the Project Proposal Evaluator***

…………………………………………………………..

Date: …………………………......