Department of Electronic & Telecommunication Engineering

Subject: Industrial Automation (UECL424)

TAE – 1

Topic: Data Analytics in Industrial Automation

Name: - Abhirat A. Bawane

Roll No: 19

Year/Semester: 4th/ 7th

Submitted To: Dr.Kapil Jajulwar.

Signature:

Title : Data Analytics In Industrial Automation

Introduction:

This paper, "Data Analytics for Industrial Process Control," discusses the challenges and solutions related to big data analysis in manufacturing, focusing on information processing and data analytics. The authors are Abdul Rauf Khan, Murat Kulahci, Henrik Schiøler, and Torben Knudsen.The paper highlights the increasing volume and precision of information generated in modern factories, which offers opportunities to understand hidden statistical dependencies in industrial processes. They propose that systematic analysis and advanced analytical methods can lead to more informed decisions.

Key Technologies:

- **Fourth Industrial Revolution/Digital Revolution:** The paper places big data analytics within the context of the fourth industrial revolution, driven by advances in Information and Communication Technology (ICT).
- **Cloud Computing:** The authors emphasize cloud computing as a solution to some big data analysis challenges, especially for small and medium-sized businesses (SMEs) due to its cost-effectiveness compared to fixed investments in computational capabilities and expertise. They define a cloud as a parallel and distributed system of interconnected and virtualized computers that are dynamically provisioned as unified computing resources based on service-level agreements.

Industrial Case Study:

• The paper uses an extended dataset from a multi-stage electronics component manufacturing process, which is approximately 10 times larger than a previous dataset.

• The objective is to predict the performance ("OK" or "NOK") of an item at the last stage (Unit) using information from previous stages (Vision, Function, and HV), which can save resources

• The dataset contains over 6 million parts, and standard PCs struggle with basic data handling due to memory limitations, necessitating the use of Apache Spark and Spark for pre-processing. Data pre-processing was done in-house due to confidentiality.

Result and Discussion:

• The genetic algorithm's convergence to the optimal solution is influenced by population size, mutation ratio, and crossover probability.

• Increased population size leads to earlier convergence and more stable results.

• The study found that beyond 5 cores, the marginal gain in speed-up or reduction in elapsed time diminishes, indicating that simply adding more resources doesn't infinitely increase computational speed.

• Larger population sizes result in a larger proportion of the parallelizable part (p), meaning parallelization is more suitable for bigger and more complex computational challenges.

Conclusion:

The paper concludes by affirming the efficacy of its proposed data mining methodology for solving categorical and combinatorial classification problems in industrial manufacturing. The iterative nature of the underlying Genetic Algorithm makes it well-suited for distributed computing environments.