

Study of Data Analysis on Machines of Pharmaceutical Manufacturing Industry

Utkarsh Rahane¹, Vishal Karande², Vrushabh Deogirikar³, Rajas Chaudhari⁴, Avinash Bagul⁵

^{1,2,3,4}B.E. Student, Dept. of Computer Engineering, NBN Sinhgad School of Engineering, Ambegaon, Pune-411041, Maharashtra, India

⁵Professor, Dept. of Computer Engineering, NBN Sinhgad School of Engineering, Ambegaon, Pune-411041, Maharashtra, India

Abstract - A checkweigher is an automatic machine used for checking the weight of packaged products. It is usually found at the terminating phase of a production line. It is used to ensure that the weight of the packaged goods is within specified limits. If any discrepancies are found, then the packages are taken out of the production line immediately. It is mostly done on fragile goods whose weight content makes a difference. Depending upon the carton size and the accuracy to be carried out, a checkweigher can weigh in excess of upto 300 items per minute. The tablet production capacity in a pharmaceutical manufacturing industry is about 15000 to 20000 tablets per hour per machine. During this process, there is a possibility of metal particles which can get infused in the contents of the tablets. For discarding such impurities in the tablets, these tablets are then passed through the pharmaceutical metal detector which discards the tablets with all possible impurities. Based upon the rejections in a specific time slot, the data would be analysed over a time span and maintenance as well as traceback activities can be carried out effortlessly. We aim to analyse the reports generated by these two machines that are used in a pharmaceutical manufacturing industry. Based on the analytics obtained, we will be determining the faults or exceptional cases occurred in the machine when working and increase their efficiency.

conveyor belts. It has a very high speed rate of weighing the packages and verifying their limits.

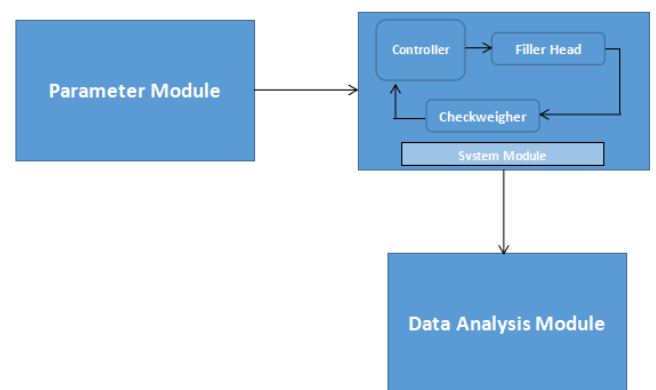


Fig -1: Checkweigher Block Diagram

A metal detector is a device that is used to detect the metal impurities present in the capsules or tablets that are being manufactured at that time. The manufactured tablets are passed through metal detector which discards the tablets with such impurities in it.

Key Words: Checkweigher, metal detector, pandas, pharmaceutical industry, Jupyter Notebook, analysis, graphical representation, Py-Qt, error rate.

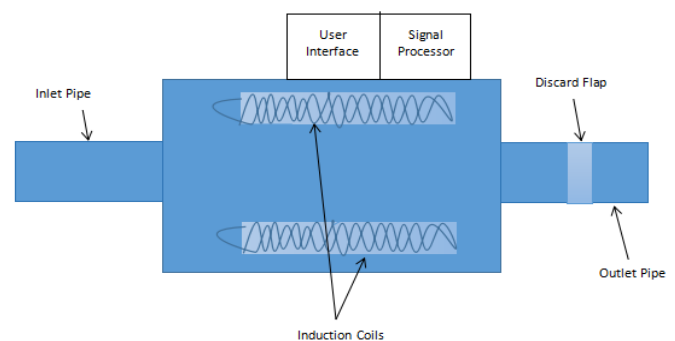


Fig -2: Metal Detector Block Diagram

1. INTRODUCTION

Data analysis plays a very important role in examining the trends and patterns in a specific domain. Pharmaceutical manufacturing industries have also been in the long run for manufacturing various sorts of tablets and medicines. The recent machines also provide the data based on their specific operation in the manufacturing process. Taking the two machines into consideration i.e. Checkweigher and Metal detector, we will be analysing the data and visualize it with various visualisation techniques.

A checkweigher is an automatic dynamic weighing machine that checks the weight of packaged commodities. It ensures that these packages are in the specified limits. If not, then these packages are discarded immediately from the

The record is maintained by these machines which can be extracted from them by inserting a flash device. Further, we can analyse this data by applying some specific libraries like numpy, pandas and matplotlib for plotting and visualisations. The same have been discussed briefly in the next section.

2. LITERATURE SURVEY

The authors Metta Ongkasuwan and Wut Sookcharoen [1] have discussed the study related to benefits of data analytics for stability and continuity of service and operation management in medical equipment industry which is used to provide insight and information for management of data generated by machines and to make decisions in managing and sustaining the production.

Ayeda Sayeed, Soo Jeon and Theodore Pribytkov [2] have thrown a light on the working of checkweigher machine. Checkweigher needs the products singulated to give the most effective results. They have proposed a IMWS system to overcome this problem caused by typical checkweigher. It enhances the capabilities of a typical checkweigher but also includes multiple hardware requirements along with an array of motion weight sensors.

Xie Xiaowen, Chen Lin, Fang Hengzhi and Chen Yongxing [3] have stated the complete pharmaceutical production process of manufacturing. The process monitoring and quality management system of modular pharmaceutical production line consists of configuration control software module, formula and parameter optimization module.

Authors Emir Slanjankic, Haris Balta, Adil Joldic, Alsa Cvitkovic, Djenan Heric, Emir Veledar [4] have emphasized on Principal Component Analysis (PCA) and Disjoint Cluster Analysis (DCA) for data reduction and summarization. DCA groups observations into clusters such that members of the same cluster have strong correlation and members from different clusters have weak correlation.

Mr. Subhashish Kumar, Dr. Namrata Dhanda and Mr. Ashutosh Pandey [5] have quoted their research on statistical analytics expertise field i.e. Data science. Use of open source library of python language for data set mining, modelling and visualization has been studied. This research has highlighted the specifications and use of Pandas, it is an open source library easy to handle data structures, algorithms and data analysis tools for python programming language.

HONG-BIN GAO and WEI-YI PANG [6] has pivoted the aim at the correction on dynamic measurement errors. Belted dynamic weight equipment is designed for continuous weighing the bulk solids or powders on conveyor belt. Measurement errors on the electronic belt conveyor scale can be divided into static measurement errors and dynamic measurement errors. To correct the dynamic measurement errors a measurement area is built and Stress Analysis of the Measurement Area is done. The method presented here is being characterized by moderate computation and high application value. Mathematical model analysis and neural network error correction method are also present. A high-accuracy dynamic weighing system has been proposed in this paper.

Joao Felipe Pimentel, Leonardo Murta, Vanessa Braganholo, and Juliana Freire [7] In this paper, the authors present a study that aims to provide insights into the reproducibility aspects of Jupyter notebook. Jupyter Notebooks are being widely used by many different communities, both in science and industry. They support the creation of literate programming documents that combine code, text, and execution results with visualizations and all sorts of rich media. This paper has three main contributions. First, it analyzes evidence of good and bad practices on the development of Jupyter Notebooks regarding reproducibility. Second, it presents a full reproducibility study that measures the reproducibility rate of notebooks (RQ7). Finally, it proposes a set of good practices that intends to minimize the criticisms and raise the reproducibility rate.

3. METHODOLOGY

NumPy is a general-purpose array processing package. It provides a high performance multidimensional array object, and tools for working with multidimensional arrays and matrices. NumPy primarily targets the CPython reference implementation of python as it is a non-optimising bytecode interpreter. This project utilises numpy for useful linear algebra and random number capabilities on the data generated by checkweigher machine.

Pandas is an open source library that allows us to perform data manipulation in Python. Pandas library is built on top of Numpy, which means Pandas needs Numpy to operate. Pandas provide an easy way to create, manipulate and wrangle the data. Pandas is also an elegant solution for time series data. Advantages of pandas are it easily handles missing data, provides an efficient way to slice the data. Panda library will be used in this project to do analysis on the data generated by the checkweigher machine. Analysis will be done for upper and lower bound of defective pieces, repetition of defects on a daily basis, weekly basis and monthly basis, analyses rate of error occurred and classified accordingly into under, best and overweight.

Matplotlib is a library for the python programming language. It has numerical mathematical extension Numpy. It has an object oriented API for embedding plots into applications by using general purpose GUI toolkit. Visual display analysis will be done with the help of this library.

The jupyter notebook is used as a platform for processing the data. It is an open source web application. It allows us to create documents that contain codes, visualization and equations.. This project utilizes jupyter notebook for numerical simulation, data analysis and for writing codes. The stated three libraries will be integrated into the Jupyter notebook which will be then used to visualise the data along with it's statistics.

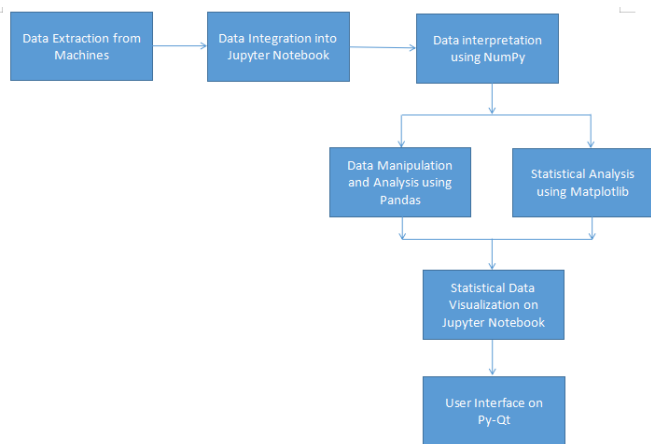


Fig -3: System Architecture

3.1 Process Flow

The complete process is explained in the following phases.

Phase 1: Data Collection

The data is continuously generated in a specific structure by the manufacturing machines. In this phase the data is collected from these machines (Checkweigher and Pharmaceutical Metal Detector) which is basically a .csv file format. This data is then fed as input data in Jupyter Notebook for further processing.

Phase 2: Environment Setup

In the second phase multiple required libraries such as NumPy, pandas and Matplotlib are integrated into the environment of Jupyter Notebook for performing various analytical techniques.

Phase 3: Data Analysis

This is the phase where statistical and logical techniques are applied systematically for analyzing the data. NumPy is used to interpret this data into required format. It is used for manipulating these numerical values to generate analysis. Further statistical analysis is done using pandas for the numerical values obtained by the data interpretation on the data. Matplotlib is used for producing graphical structures based on this data.

Phase 4: Data Visualization and Interface

In this phase visualization is done in Jupyter notebook. Further, the front end interface is done with the help of functionalities provided by Py-Qt. The results of analysis phase are displayed in the form of various graphical structures and reports are generated.

3. CONCLUSION

Hence, by using various analytical techniques on the data of checkweigher and metal detector machines in pharmaceutical industry we have generated some graphically represented results and reports which will be helpful in determining faults or exceptional cases occurred while working and so resulting in increasing efficiency of the machines.

REFERENCES

- 1) Metta Ongkasuwan and Wut Sookcharoen "Data Analytics for Service and Operation Management Improvement in Medical Equipment Industry" in 2018 5th International Conference on Business and Industrial Research (ICBIR), Bangkok, Thailand.
- 2) Ayeda Sayeed, Soo Jeon and Theodore Pribytkov, "In-Motion Weight Sensor Array for Dynamic Weighing of Non-Singulated Objects"
- 3) Xie Xiaowen, Chen Lin and Fang Hengzhi, Chen Yongxing, "Research on Process Supervision and Quality Control System of Modular Pharmaceutical Production Line".
- 4) The authors Emir Slanjankic, Haris Balta, Adil Joldic, Alsa Cvitkovic, Djenan Heric and Emir Veledar, "Data Mining Techniques and SAS as a tool for graphical presentation of Principal Components Analysis and Disjoint Cluster Analysis results".
- 5) Mr. Subhashish Kumar, Dr. Namrata Dhanda, Mr. Ashutosh Pandey, "Data Science – Cosmic Infoset Mining, Modeling and Visualization", 2018 International Conference on Computational and Characterization Techniques in Engineering & Sciences (CCTES) Integral University, Lucknow, India, Sep 14-15, 2018.
- 6) HONG-BIN GAO and WEI-YI PANG, "A HIGH-ACCURACY DYNAMIC WEIGHING SYSTEM BASED ON SINGLE-IDLER CONVEYOR BELT", Proceeding of the Eighth International Conference on Machine Learning and Cybernetics, Baoding, 12-15 July 2009.
- 7) Authors Joao Felipe Pimentel, Leonardo Murta, Vanessa Braganholo, and Juliana Freire "A Large-scale Study about Quality and Reproducibility of Jupyter Notebooks", 2019 IEEE/ACM 16th International Conference on Mining Software Repositories (MSR).