International Journal of Future Generation Communication and Networking
Vol. 13, No. 3s, (2020), pp. 1728–1736

Data analysis on machines of pharmaceutical manufacturing Industry

Utkarsh Rahane[1], Vishal Karande[2], Vrushabh Deogirikar[3], Rajas Chaudhari[4], Prof. Avinash Bagul[5]

[1,2,3,4]B.E. Student, Dept. of Computer Engineering, NBN Sinhgad School of Engineering, Ambegoan, Pune- 411041, Maharashtra, India
[5]Professor, Dept. of Computer Engineering, NBN Sinhgad School Of Engineering, Ambegoan, Pune- 411041, Maharashtra, India

Abstract

Pharmaceutical manufacturing industry involves a variety of machines for various purpose like production, quality check, packaging and many such. Checkweigher and metal detector are machines being used in this industry. Checkweigher is a machine used for checking the weight of packaged products. It is used to verify whether the weight of the packaged goods is in the predefined limits or not. It can also categorize the products as per their weight in specific range. If any discrepancies are found then the packages are discarded from the production line immediately. It is mostly done on fragile goods whose weight content makes a difference. A checkweigher can weigh upto 300 items per minute. In a pharmaceutical manufacturing industry about 15000 to 20000 tablets per machine per hour are produced. During this process, there is a possibility that metal particles may get infused in the contents of the tablets. For checking and discarding such impured tablets, these are passed through the pharmaceutical metal detector which discards the tablets with all possible impurities. We aim to analyse the data generated by both these machines and generate reports that can be used in a pharmaceutical manufacturing industry for traceback activities. Based on this analysis and graphs generated, we will be determining the faults or exceptional cases occurred in the machine or production line when working which will provide assistance in increasing efficiency.

Keywords: *Checkweigher, metal detector, pandas, pharmaceutical industry, dash, analysis, graphical representation, anaconda, error rate.*

I. Introduction

The pharmaceutical manufacturing industry produces pharmaceutical drugs for use as medications to patients. Pharmaceutical manufacturing continues to evolve with use of new emerging scientific technologies. Effective use of current science and engineering principles throughout the life cycle of a product can improve the efficiencies of both the manufacturing processes and quality checking.

There are various machines used in this manufacturing process. Checkweigher is used for validating whether products are in the specified weight category if not else discardes the ones which are overweight or underweight. Metal detector is another machine used for discarding those tablets having even a little metal particle in them. The recent machines also provide the data and graphs based on their specific operation in the manufacturing process.

ISSN: 2233-7857 IJFGCN

Data analysis field plays a very vital role in examining the trends and patterns in a specific domain using the data. The data created by these machines can be extracted from them using a flash device. We can get the working time, machine status, rejection count in metal detector data and time, distribution according to weight categories, discarded count in checkweigher data. In this paper, we propose a system which will visualize the data of these machines using various graphs and will help in making further predictions regarding machine health status and products manufactured.

Our system uses dash interface along with anaconda navigator. The data produced from the machines will act as an input to our model. Various python libraries have been used in our system for analysis and visualization purpose. We will be getting various graphs on the interface along with reports for effective comparison with previous results. Run time analysis will be there along with feature of using a .csv or excel sheet as an input .

The remainder of the paper is sorted in the following manner. Section II gives the Motivation and Need for our project. Section III discusses the Literature Survey which demonstrates how data analysis is being used in development of pharmaceutical manufacturing industry. Section IV makes you acquainted with the existing framework and our proposed framework. Segment V elaborates on our proposed framework in detail covering tools and libraries of our site. In section VI results of our experiment have been shown. And finally, we conclude our topic in section VII.

II. Motivation

Pharmaceutical manufacturing industries have been in a long run for manufacturing various sorts of tablets and medicines. So, these machines need maintenance on time to increase their lifetime and to work efficiently. Some of the machines have a system for providing data analysis during functioning only. We chose this topic to develop a system which can analyze previous records as well along with comparison analysis which would help in decision making regarding machine health status and produced goods. This way we can help in increasing machine lifetime by monitoring its health and performance continuously.

Need

There are a lot of patients who rely on the medications prescribed by their doctors. All of these medical byproducts once manufactured, are passed through the validation machines. So as a verification check, these machine prove to be quite useful in the process. Therefore they need to be in proper condition time to time, for this we can utilize various analytical techniques and methods on the reports of these machines.

III. Literature Survey

[1] "Data Analytics for Service and Operation Management Improvement in Medical Equipment Industry"

The authors Metta Ongkasuwan and Wut Sookcharoen have the benefits of data analytics for stability and continuity of service and operation management in medical equipment industry which are explained in

ISSN: 2233-7857 IJFGCN

this paper. Data analytics field can provide insight and information for management of data generated by machines and to make decisions in managing and sustaining the production.

[2] "In-Motion Weight Sensor Array for Dynamic Weighing of Non-Singulated Objects"

Authors Ayeda Sayeed, Soo Jeon and Theodore Pribytkov have explained the working of checkweigher machine. Checkweigher can give the most effective results when the products are singulated. An IMWS system is being proposed to overcome this problem caused by typical checkweigher. It includes multiple hardware requirements along with an array of motion weight sensors but enhances the capabilities of a typical checkweigher.

[3] "Research on Process Supervision and Quality Control System of Modular Pharmaceutical Production Line"

Authors Xie Xiaowen, Chen Lin, Fang Hengzhi and Chen Yongxing have studied the complete pharmaceutical production process of manufacturing and have explained it in this paper. The quality management system and process monitoring of modular pharmaceutical production line consists of formula and parameter optimization module, configuration control software module.

[4] "Data Mining Techniques and SAS as a tool for graphical presentation of Principal Components Analysis and Disjoint Cluster Analysis results"

Authors Emir Slanjankic, Haris Balta, Adil Joldic, Alsa Cvitkovic, Djenan Heric, Emir Veledar have focused 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.

[5] "Data Science – Cosmic Infoset Mining, Modeling and Visualization"

The authors Mr. Subhashish Kumar, Dr. Namrata Dhanda and Mr. Ashutosh Pandey 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.

[6] "A High-Accuracy Dynamic Weighing System Based On Single-Idler Conveyor Belt"

The authors Hong-Bin Gao and Wei-Yi Pang 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 and dynamic errors. To correct the dynamic 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.

ISSN: 2233-7857 IJFGCN

[7] "A Large-scale Study about Quality and Reproducibility of Jupyter Notebooks"

The authors Joao Felipe Pimentel, Leonardo Murta, Vanessa Braganholo and Juliana Freire 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.

IV. System Architecture

1) Existing System

The machines that we are considering i.e. metal detector and checkweigher both are validation machines used checking the quality of products manufactured before packaging them.

a) Metal Detector.

A metal detector is a device that is used to detect the metal impurities present in tablets that are being manufactured. The manufactured tablets are passed through metal detector which discards the tablets with any minor metal impurities in it. They contain a couple of induction coils which are used to interact with metallic elements in the medicines. The metal particle if present creates a magnetic field within the detector coil which in turn opens the rejection flap by means of solenoid through which the tablets fall down in the rejected block. Raspberry Pi is used for interface on a screen which is mounted on the machine. On the screen you can access the data which is in a tabular format containing date of functioning, batch no. of tablets, total no. of hours the machine worked, status of machine (safe zone, acceptable zone, danger zone) and the rejection rate in each batch. It also has a proper login so that the data can be accessed by only authenticated users. The records are also available in .csv file format.

b) Checkweigher

A checkweigher is an automatic dynamic weighing machine that checks the weight of packaged items. It ensures that these packages are in the specified weight limits. If not then these packages are discarded immediately from the conveyor belt. It has a very high speed of weighing the packages and verifying whether they fit in the range or are underweight, overweight. It usually consists of four units: weighing conveyor unit, infeed conveyor unit, control unit and base unit allowing fast and simple assembly. Output data contains : date, time, group product number, weight data and judgments. Various summary data such as histograms, R charts, total summaries, and others are available. Operational history and weighing history are also there.

ISSN: 2233-7857 IJFGCN

Disadvantages:

1) Metal detector does not have any feature for getting operational summary and visualizations.
2) The charts which are available could only summarize the operations and results.
3) Manual work is needed to find out patterns in the results.
4) Can not help in any making any predictions regarding machine status.

2) Proposed System

In Proposed system we are making use of Anaconda platform for importing various libraries and for integrating various tools. In Proposed system, we intend to visualize the data generated from these machines using various graphs like pie chart, line cart, etc. which will be helpful in summarizing the results on a web based application. Also, reports will be generated with some statistical analysis in it providing some assistance in making further predictions on the product manufactured. As per needs admin can select a date range as well to get analysis of data recorded between those dates only. The reports and analysis will help in taking some decisions on whether the machine needs some maintenance or not. Also fluctuations among the daily rejection rate can also be seen which may help in concluding that whether there has been some issue in the batch manufactured or the health of machine. This system will make it possible to view the analysis of multiple validation machines on a single portal also is having web interface so can be accessed from anywhere.

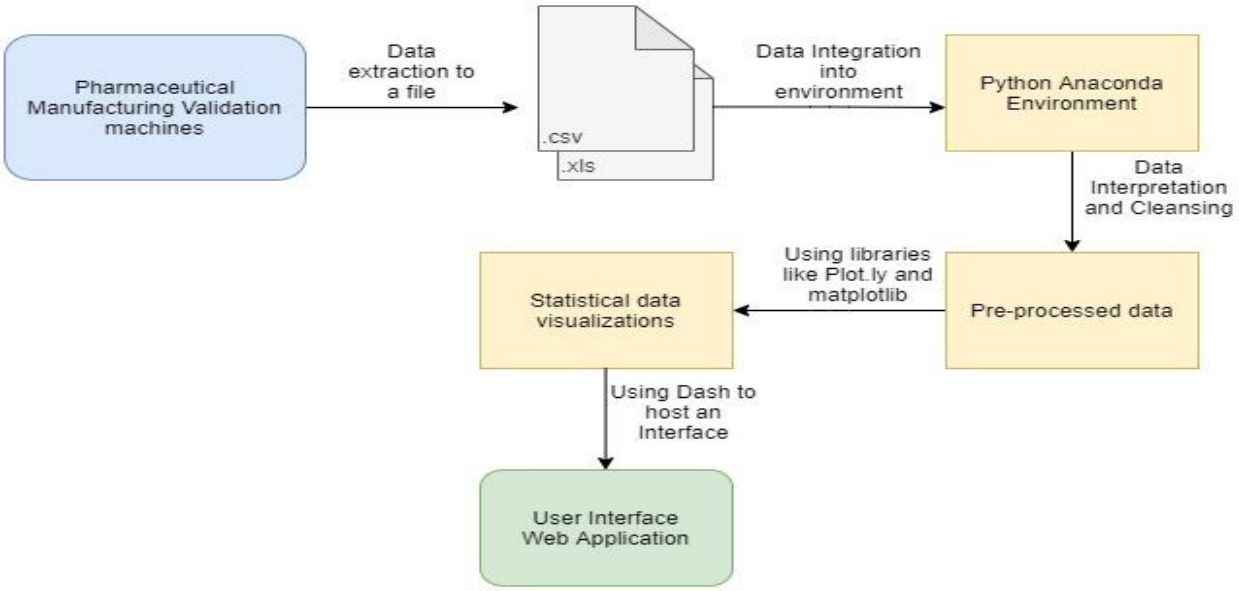


Fig 5.1 Proposed System Architecture

ISSN: 2233-7857 IJFGCN

Advantages:

1) Helps the administrator in making analysis and taking some decisions.
2) System is platform independent as it is web based application.
3) Accepts data in both .csv as well as .xls format**.**

V. Tools Used In Proposed System

Anaconda :-

Anaconda is an open source platform for scientific computing. It provides a lot of preinstalled libraries for data science with python.

We are using Spyder 3.3.6 in Anaconda as the basic editor and compiler for python application alongwith various useful and related libraries. Spyder integrates with many packages like NumPy, pandas, Matplotlib, IPython, SciPy and many such. It is a powerful python IDE which provides advanced and interactive tsting and debugging fetures. We are using it for python application along with other useful libraries.

Jupyter Notebook :-

The jupyter notebook is used as a platform for processing data. It is an open source web application which allows us to create documents that contain codes, visualization and equations. One major feature of the jupyter notebook is the ability to display plots that are the output of running code cells. The matplotlib plotting library provides this functionality.

Some libraries that we will be using are :-

1) Dash :-
   Dash is a user interface library for creating web applications. It is used for data analysis, data exploration, interactive visualization, modelling and reporting purpose in python. It provides a Python interface to a rich set of interactive web-based components. Dynamic visualizations using plotly are also displayed on dash interface.

2) Matplotlib :-
   Matplotlib is a python library. It has an object oriented API for embedding plots into applications by using general purpose GUI toolkit. Visual display analysis can be done with this library. It is used for creating interactive, static and animated visualizations in python.

ISSN: 2233-7857 IJFGCN

3) Plotly
Plotly is a python based library used for making various statistical visualizations. The most important feature of plotly is that it allows to plot dynamic graphs. It makes interactive graphs which can be dynamically displayed on the dash interface.

4) Pandas :-
Pandas is an open source library that performs data manipulation in Python. It provide an easy way to create and manipulate the data. Advantages of pandas are it easily handles missing data and provides an efficient way to slice the data. We will use this library for performing analysis on the data generated by the checkweigher machine. Analysis will be done for max and min weight limit 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.

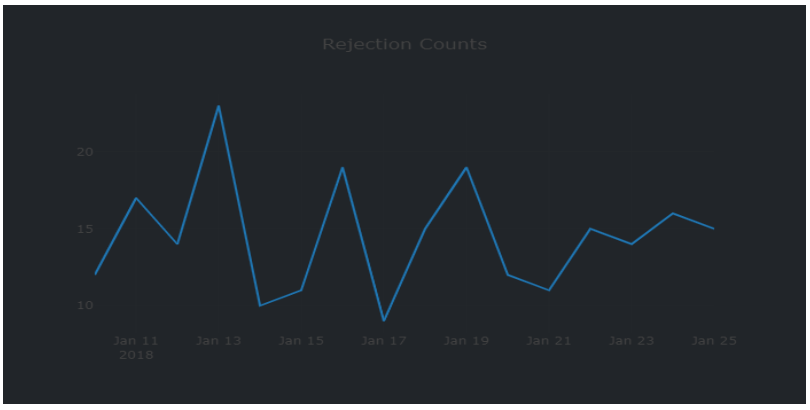
5) NumPy :-
NumPy is a python library having high-level mathematical functions to operate on large multi-dimensional arrays and matrices. This project uses numpy for useful linear algebra and random number capabilities on the data generated by checkweigher machine.

VI. RESULT

1) Result

a) Line Chart Visialization :-

On the interface we will be getting various graphs being plotted for analysis purpose. This is a line chart showing the variation in daily rejection count of metal detector irrespective of the machine status.

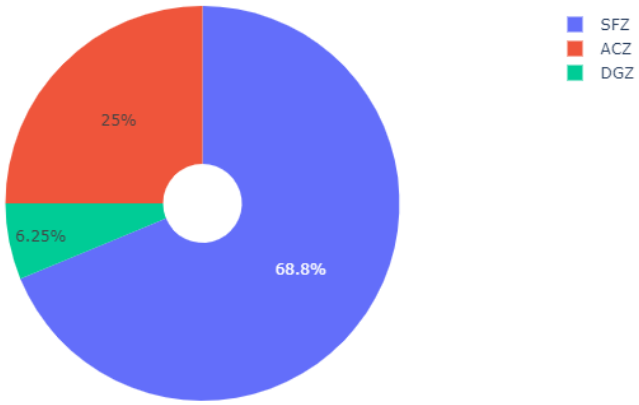


b) Pie Chart visualization :-

This is a pie chart showing the percentage wise distribution of working time of machine on the basis of machine status i.e. Safe zone, Acceptable zone, Danger zone. By this visualization we can see that the machine has worked in the danger zone for a very less time so we can conclude

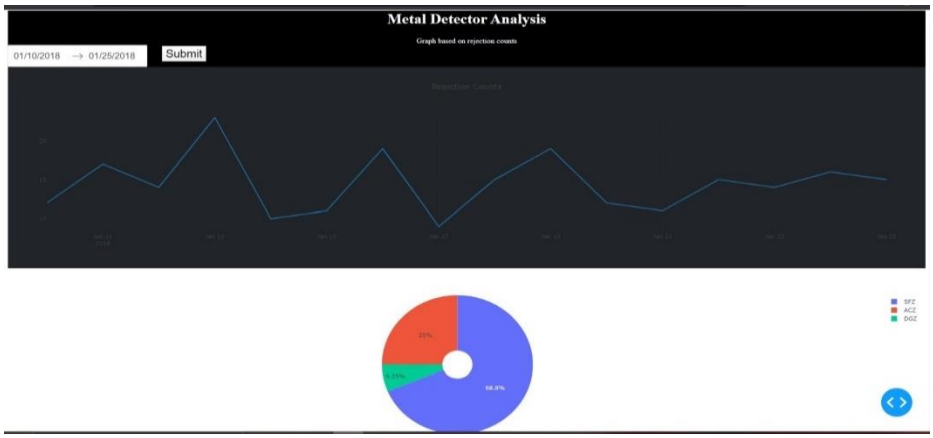
ISSN: 2233-7857 IJFGCN

that.



c) Overall interface :-

This is how the overall interface will look like. We also have a date selector on top which helps us select a date range to get the visualizations on the data of that dates only.



VII. Future Work & Conclusion

Future Work

As we have developed this project at college level we are using data for some limited time period. But our system is feasible for any date range for which data is available. In future if there is need we can use this system for other machines as well by changing some of the parameters as per requirements. More graphs for visualization can also be plottedby some simple modifications.

Conclusion

In this paper, we proposed a web based system for analysis and visualization of data produced

ISSN: 2233-7857 IJFGCN

by pharmaceutical machines metal detector and checkweigher. For this purpose, we built a python based system on which data sheets can be analysed and various graphs can generated along with reports. This system proves out to be very useful in making some decisions regarding machine health status by analyzing its performance and some decisions regarding the quality of products manufactured. We have made this system for only these 2 machines right now but it can be used for other machines as well by just changing some parameters.

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).

ISSN: 2233-7857 IJFGCN