***Report on***

***Data mining in pharmaceutical industry***

**Heena Dawani - 27**

**Navin Dhamriyani - 28**

**Riya Karia - 44**

**Introduction :**

Data Mining, the process of extracting information from large data sets through the use of algorithms and techniques drawn from the field of Statistics, Machine Learning and Database Management Systems.

“**Mining**” means to find something that already exists.Therefore, data mining can be defined as a process of identifying hidden patterns and relationships, and trends within data.

The amount of data getting generated in any sector at present is enormous. The information flow in the pharma industry is huge. Pharma firms are progressing into increased technology-enabled products and services. Data mining, which is knowledge discovery from large sets of data, helps pharma firms to discover patterns in improving the quality of drug discovery and delivery methods. Applications such as sales force planning and direct marketing to doctors and consumers are employing many new data mining techniques.

**Traditional methods often involves :**

1. Manual work
2. Interpretation of data.

**Six important steps in the Data Mining process as :**

1. Problem Definition.
2. Knowledge acquisition.
3. Data selection.
4. Data Preprocessing.
5. Analysis and Interpretation.
6. Reporting and Use.

**Relevant data sources for the pharma industry are** **:**

1. **Clinical data** (Chemical composition, demographics, pharmaceutical data, medical treatments, length of stay).
2. **Administrative data** (staff skills, overtime, nursing care hours, staff sick leave).
3. **Financial data** (treatment costs, drug costs, staff salaries, accounting, cost-effectiveness studies).

**Predictions :**

1. Predicting consumer behavior.
2. Predicting the likelihood of success in a drug adoption process.
3. Predicting the percentage accuracy in performance of a drug.
4. Classifying the historical health records.
5. Prediction of what type of drugs most likely to be retained, most likely to be left, most likely to transform their composition.
6. Predicting pharma product behavior and attitude.
7. Predicting demand projections by seasonal variations.
8. Identifying the best profile for different drugs.
9. Classify trends of movements through the organization for successful/unsuccessful patient historical records.
10. Categorization of drugs with respect to diseases and patients.

**Clustering**(grouping of data) **:**

1. It is a method by which similar records are grouped together.
2. Clustering is usually used to mean segmentation.
3. An organization can take the hierarchy of classes that group similar events.
4. Using clustering, patients can be grouped based on age, name, diseases, criticality, locality, etc. This will help the pharmaceutical companies to match the demand and supply ratio and the various factors affecting the sales .
5. In business, clustering helps identify groups of similarities, characterize customer groups based on purchasing patterns, etc.

**Use Case :**

Mining can help us to measure the chemical activity of the molecule on specific disease say tuberculosis and find out which part of the molecule is causing the action.This way we can combine a vast number of molecules forming a super molecule with only the specific part of the molecule which is responsible for the action and inhibiting the other parts.This would greatly reduce the adverse effects associated with drug actions.

**Applications :**

**1) Clinical data analysis** – clinical data analysis evaluates and streamlines from large amount of information. Data mining helps to see trends, irregularity, and risk during product development and launch.

**2) Marketing and sales analysis** –the identification of the most profitable product and allocation of marketing funds. Data mining here helps to examine consumer behavior in terms of prescription renewal and product purchases.

**3) Customer analysis** – using data mining one can develop more targeted customer profiles that focus not only on products, but also on the ability to pay for them by analyzing historical health trends in combination with demographics.

4) **Target physicians** who have high prescription rates of a certain drug or treatment with new drug information that treat complementary symptoms or conditions.

**Conclusion :**

Data mining is process of utilizing the scientific methods to manage the pharmaceutical industry, it has varied usage felt on new drug discovery, track record on clinical trials, composing the new molecule feature with other features. Predicting the behavior of drugs and predicting the demand, managing the demand supply ratio. In future it will be combined with the recent trends like Internet of Things, cloud computing and big data analytics in order to increase the efficiency of the process.