Johnson & Johnson Vision Final Round

Ask #1: Case Study

1. Introduction

In today's world, every industry is moving towards Automated Manufacturing (use of intelligent machines) and we have witnessed a surge in the application of machine learning in the manufacturing industry over the last 2 decades. With the help of machine learning, these manufacturing industries have been able to augment their capacity and have been able to make the process more efficient by reducing the errors.

2. Use of Machine Learning Models in Automated Manufacturing Industry (J&J Vision)

- We can use ML models in J&J Vision provided technology cataract surgery to automatically detect cataracts from color images by applying **Deep Neural Network** technique to achieve high accuracy result by extracting the features like uniformity, intensity and standard deviation. With **Deep Neural Network**, we can also look for visual clues in the photographs of patient's retina that will show the presence of the medical condition. Neural Network image recognition system is very useful ML model for healthcare diagnostics.
- Machine learning model like **Apriori algorithm/ k-means Clustering** is helpful in determining early stage drug discovery with the help of tons of data by removing complicated biological system. Also, in personalized medicine, we can use patient data to predict responses for the particular treatment by using **Multivariate Linear Regression (MLR)**.
- Using **Ensemble ML algorithm,** we can do automated detection of diabetic retinopathy (DR) in retinal images. **Ensemble Methods** are combination of distinct learning algorithms like adaBoost, Random Forest, SVM and uses DR features to classify whether an image has diabetic retinopathy or not.
- As we know Manufacturing Sector collects huge amount of data from sensors attached to every aspect of production line during the growth. ML model **Support Vector Machine (SVM)** can be used to analyze this kind of dataset to find out the hidden trends and patterns as it gives best accuracy & more efficiency for non-linear complex datasets. Also, it's always good to use **Principal Component Analysis (PCA)** for reducing feature dimension in case of large complicated manufacturing datasets.
- ML classification techniques **Naïve Bayes'** & **Logistic Regression** are also very useful in driving collaborative robots proof in manufacturing industries to learn observation of production line & data streams which can optimize the production process to minimize production costs & to increase the speed of production cycle.

3. Conclusion

In this era of machine learning algorithms and with the increasing availability of data & computing power, the applications for machine learning, especially in automated manufacturing industry, will increase further at a rapid pace. Though supervised algorithms are used in most application, unsupervised and Reinforcements learning methods are also increasing with the increase in data. To conclude, I will say Machine learning is already a powerful tool for many applications in the manufacturing industry and its importance will increase further in the future and the industries not embracing this powerful tool will be left far behind in this race.

References:-

- http://ieeexplore.ieee.org/document/7848208/
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5082593/

Ask #2: Sample Dataset Analysis

1) What is your analysis procedure?

Mentioning step by step my procedure of analyzing the Lens Quality datasets:-

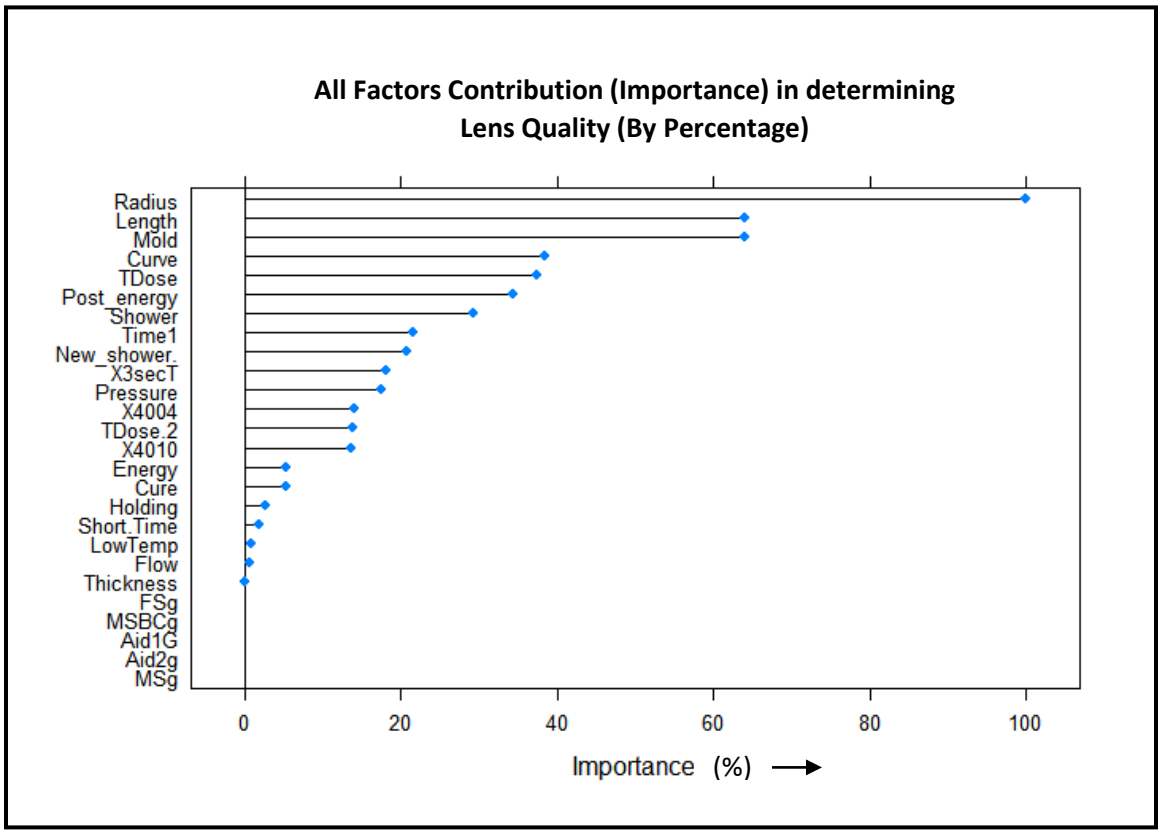
- *Step 1*:- To understand the task properly and to find out the answers: What's the requirement? What type of data is available in the dataset? What is the response variable and what are the predictors?

- *Step 2*:- Once I recognized the type of data and the problem, I moved in to a *Data Cleaning* part which focuses on removal or transformation of missing data, removal of "zero-optional" variables (numeric & nominal), Identification & removal of correlated predictors (numeric).

- *Step 3*:- Now, comes the *Exploratory Analysis* part that focuses on model selection, model training, and finally model evaluation.

- *Step 4*:- The last step is basically a proper *Interpretation & Visualization* based on our results that comes after proper model evaluation.

2) Which factors are most important in determining lens quality?



As we can see in above bar chart, most important factors (Top 10) in determining the lens quality are :- Radius (100%), Length (64.08%), Mold (64.08%), Curve (38.48%), Tdose (37.37%), Post Energy (34.40%), Shower (29.25%), Time1 (21.46%), New Shower (20.73%), 3secT (18.09%).

- Radius is major feature to determine the lens quality following with Length and Fold.
- Few of the factors that are not contributing totally are FSg, MSBCg, Aid1G, Aid2g, and Msg.
- None of the process setting factors except Pressure is contributing more than 5%.

3) Which tools did you use for your analysis?

Here, I choose 'R' over python due to its amazing built in libraries and packaging, however, we can do this analysis easily in both python & R by using several ML models.

References:-

- https://www.rdocumentation.org/packages/caret/versions/6.0-78/topics/train
- https://cran.r-project.org/web/packages/caret/vignettes/caret.pdf

Things I tried but I failed:-

- In data challenge, I tried to find out the methods of automatic change in process setting factors so that we can check if by doing so setting features are able to contribute in determining the lens quality. For this, I learned wrapper methods and R functions (Boruta), but these ways are not helpful here; I tried to achieve this in Python also.
- In case study, I wanted to add Machine Learning models that has applications in J&J vision product ACUVUE brand contact lenses, but due to lack of knowledge of product I am not sure about the performance of particular model in analyzing the product. For this, I read about this product and few IEEE papers to know more about the product & lenses.

Things I will do for Future work:-

- I will try to read more about the J&J vision product in depth and will try to check more on Python & R to deal with the setting factors.

Personal comment: I enjoyed this Case study and Data analysis a lot. It was fun to working on real world dataset and to find out the Machine learning application from J&J vision perspective. I hope I answered the questions. I tried to make it simpler. For the case study, I read white papers & ieee papers to write best ML model application in automated manufacturing industry. And, for data challenge, I checked the importance of factors by the use of two models & picked the one which has high R-squared vale, accuracy and low MSE.