As There is tremendous growth in internet users in few past years leading to a completely new environment for customers, stock holders and manufacturers to exchange ideas and feedback about the products and services being available in the market or those to be introduced.

But however we have those old traditional method like banners behind the roads, painting on the wall, distributing pamphlet are being old school practice for promotion kind of things, it might works good for some of brands, and we can’t completely ignore those practice but-but if we talk about the brands which mainly target customers from young generation, mean to say product involving costly items like cosmetics, expensive clothes, watches, cell phones, electronic gadgets, online world, etc. These are the things whose customers are so broad minded, they are enough connected to social media that they get more information, review about their choice of products. Than any other sources.

Companies soon realized the potential of using Internet-based social networks to influence customers, incorporating social media marketing communication in their strategies for influencing their businesses. A system able to predict the impact of individual published posts can provide a valuable advantage when deciding to communicate through social media, tailoring the promotion of products and services. Advertising managers could make judged decisions on the receptiveness of the posts published earlier, thus aligning strategies toward optimizing the impact of posts, benefiting from the predictions made. Also, it has been shown that social media publications are highly related to brand building.

**Coming to algorithm used for building the model:-**

We have used have the dataset provided by UCI hub as sample test case for building the model.

And we tested the model on our club pages of our college.

For real model uses the page insight information of all the action that took place on page and previous posts, using this as our training dataset for our upcoming page action we create the best result from it.

Different data mining technique used such as support vector regression , polynomial regression and somewhere linear regression too used for prediction. We adopted **the Shapiro–Wilk test** to assess if each of the output columns for the features to be predicted followed a normal distribution

After finalizing the threshold result we have used maximum likelihood estimation at the regression discontinuity of our prediction to gain more accuracy on Result.

Regression discontinuity research designs exploit the fact that some rules are quite arbitrary and therefore provide good quasi-experiments when we compare (elements, people or cities, rms, countries...) who are just aﬀected by the rule with people who are just not aﬀected by any fixed rule. There are two types of design is proposed for regression discontinuity. First is sharp regression discontinuity that is a deterministic function of a covariate X. RD captures the causal eﬀect by distinguishing the nonlinear and discontinuous function, (Xi ≥ c) from the smooth function f(Xi(2nd type of RD). In Sharp RD designs we exploit that treatment status (other factors which is not included in input features) is a deterministic and discontinuous function of a covariate Xi.

Di = (1, if Xi ≥ c

(0, otherwise

Where, c is known as threshold or cutoﬀ point. Once, we know Xi, we compute Di that is shown in ﬁgure 1 having the cutoﬀ point is 0.

Attributes like month, day, time of post, (e.g. we can’t advertise about winter clothes in summer even if we add attractive offer on it won’t benefit the brand)

We used polynomial regression as every month, day hour has different dependency. Using this the average result created from dataset is that daytime approximately 6am-3pm social media is most active.it is illusion that during night time most of people stay active on it.

Any particular Day doesn’t have extraordinary effect on promotion.

Coming to number of page like here we discovered that to make the page a remarkable and to gain a good lifetime post consumption the threshold is too reach 0.9 million likes. And to verify this we used maximum likelihood estimation at left hand limit and right hand limit with delta difference of 1000 to 20000 likes and the result is like below 0.9 million likes page is under care condition and above 0.9 million likes page is now in safe condition for tolerating some malpractice too.

We also added likes on post, reaction on post, comments, share on post as independent variable for dependant variable “lifetime post consumptions” for building the model.

Adding on type of post content (like only some link, or photos or any videos, or something any trending) this all stand very likely and too get a clear prediction here we used support vector regression of higher degree along with maximum likelihood estimation.

Creating all the result together we reached a good level of accuracy with overall error of 22.11%

Which is even better than another such paper published which has overall error of 26.9%.the reason we reached such accuracy because previous paper was totally based on linear, polynomial and multiple regression ,whereas we have used support vector regression too along with maximum likelihood estimation at regression discontinuity.

Paper was completed and verified by the mentor and it is sent for publication at international……..

Terms— support vector machine, support vector regression, **the Shapiro–Wilk test, why the Shapiro–Wilk test, maximum likelihood estimation , regression discontinuity.**