

Signature Assignment: Week 6

ALY6120 – Leadership in Analytics

Fall Quarter – 2019

Instructor – Christopher Waller



Monisha Rajendran

NUID – 001064669

Assigned On – 16th September 2019

Submitted On – 28th October 2019

Introduction:

As a CEO of a research data management organization which focus on servicing Lab applications in Drug discovery domain, I see a lot of dissatisfactions with my clients which is resulting in customer churn where the agreed SOW contracts are not renewed. Performing a high-level analysis, I found that the services which we are providing are all built in-house from respective client side. This is a potential threat to my company's current revenue and growth. This is a hypothetical situation that I would love to own and address it from the learnings of this course as a leader using CRISP model.

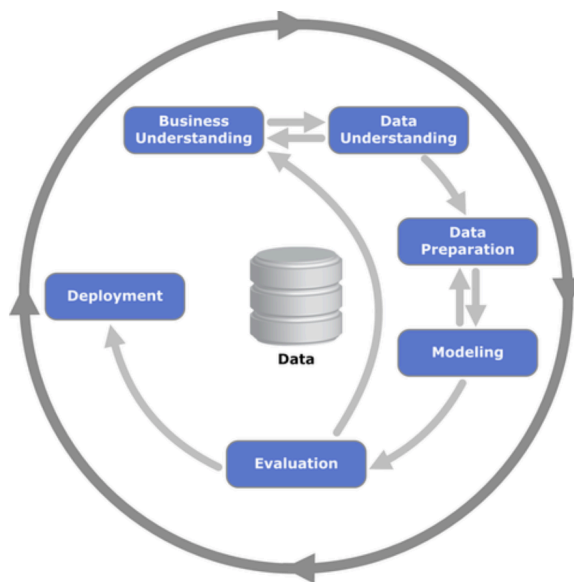


Figure 1.0 [Phases of CRISP-DM Process Model]

As a leader, I always believe team as an asset. I could easily establish a rapport with colleagues. As I have prior experiences in working with time critical projects with more user base, I feel to take up this challenging role of leading and driving these initiatives. On that note, I need to gain more hands-on experiences in following data integrity and documentation procedures as these are crucial elements in final deployment. Also, I am targeting to learn the art of listening which measures my success factor for value driven feedback approaches.

Understanding the business problem:

Here are few business objectives for which I would like to implement analytics for deriving solutions:

1) Focus on uncovering business insights connecting different departments inside the organization. 2) Understand new areas of expansion that will be the right step for venturing into a new department within the organization. 3) Using the available data from existing clients, need to come up with insights for making changes in the organization into its new venturing that will increase ROI. To achieve that below are few data sources from which I am planning to collect pre-requisites for mining. 1) Collect employee's input and know the pit falls which is creating a gap in overall business process. 2) Define variables and collect global data to understand the trend in drug discovery and development process. 3) Collect customer data to identify and focus on customer's interest.

Understanding the data:

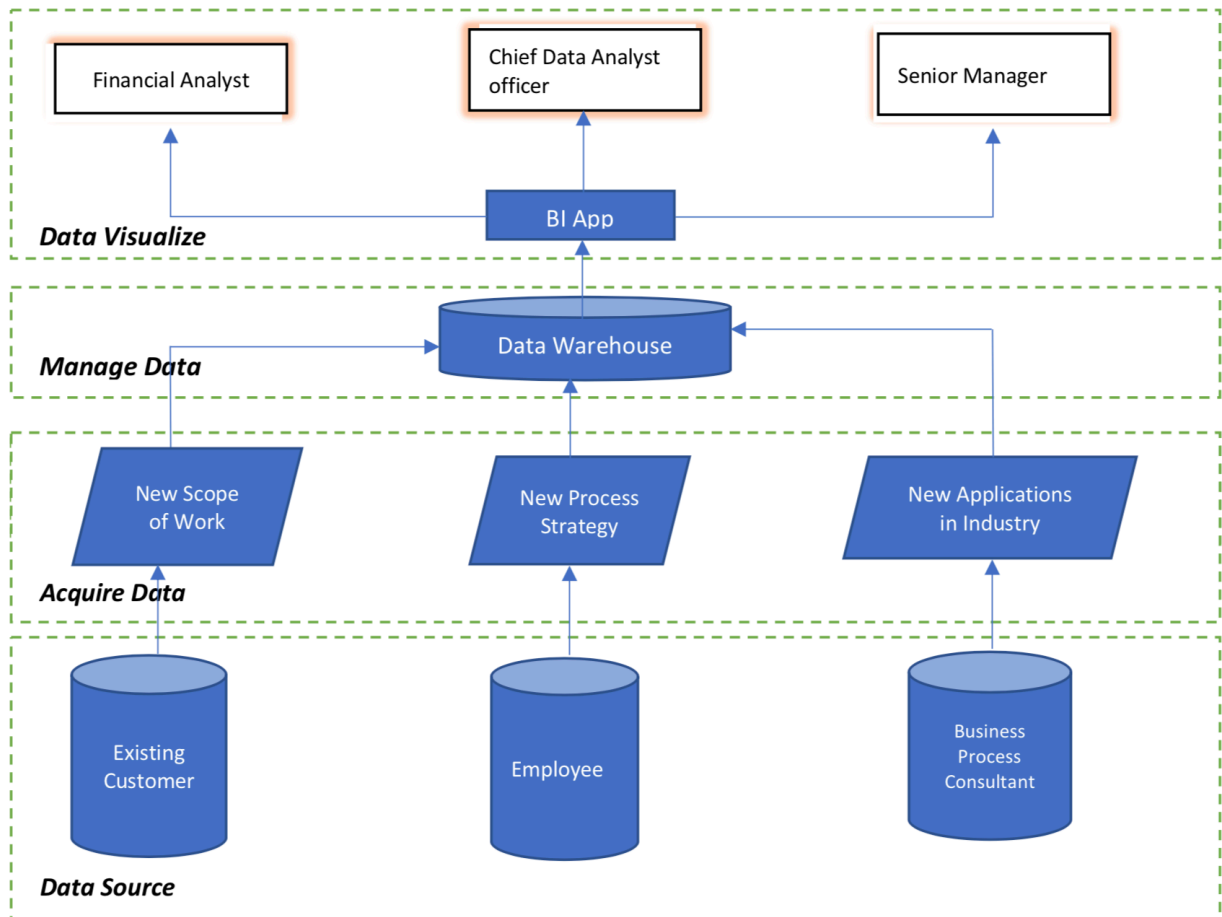
I think this is the time to make primary objectives addressing each of the business problem with the business goal clearer. Below are the purposes for addressing above problems.

1. ↑ Business process strategies 2. ↑ Business area expansion 3. ↑ Revenue of the organization

1. Identify 3 new process strategies 2. Categorize two new Application that is used by many customers in this domain 3. Discover two new scope of work from at least three existing clients.

User roles	Use case	User stories
QA Representative	Identify team members	Determine employee inputs
Senior Manager	- Validates insights - Performs decision making	- Determine new process strategies - Finalize business area expansion actions
Business Process Consultant	Explore new applications	Reports the list of new applications
Strategy Analyst	- Prepare questionnaires - Analyze market scope of applications	Determine business continuity by knowing customer scope
HR Manager	Analyze org. skill metric	Determine the availability of potential resources

Team Lead	Meet customer to get answers for questionnaire	Complete questionnaires
Data Engineer	Analyze raw data	- Determine employee survey insights - Determine customer survey insights
ML developers	Create predictive model	Determine customer behavior
Chief Data Analyst officer (CDAO)	Coordinate with teams for finalizing the overall strategies	Deployment of goals in action
Financial Analyst	Analyze project profitability	Determine profit outcomes

Architecture Diagram:

Data preparation

Business process strategy improvements: A potential threat to my company's revenue and growth lies in addressing gaps in current business process strategies. Until now, departments in my organization acts in silos and I believe dealing with these disconnected teams are the key. On that note, I am proposing the questionnaires to get the data for the below attributes internally from a group of 20 people within the organization across leads of Product development; Product design; Validation; Applied Research; Customer Service departments. 1. The number of projects handled within the last 2 years which involved cross-department collaboration 2. No. of projects that did not meet timelines due to miscommunication 3. Projects that had misaligned goals and priorities. 4. % of trust and confidence in handling cross-department work. 5. Rework done without awareness of the work already done in another department. 6. Team's understanding on expectations of work to be done by them when working in cross-department project. 7. Time spent on understanding project scope due to differences in opinion. Answers are pre-processed and parameters with more Information Gain (IG) is understood for formulating strategies. (*QA representative*) - Provides answers detailed in questionnaire; (*Data Engineer*) - Pre-processes the data formats and identifies scenario trend; (*Senior Manager*) – Formulates new business Standard Operating Procedures and principles that aid cross-functional business process strategies. I believe predicting insights from these actions would lead the way. Business area expansion:

(*Business Process Consultant*) 1. List scope for new applications in the domain that our organization can venture into (Scraping through website, attend conferences, check Gartner magic quadrant) – External sources of data 2. Categorize the pros and cons (*Strategy Analyst*) 1. Performs market trend analysis and comes up with final two applications which he discusses with *Senior Manager* and chooses final one upon evaluating data relevancy and accuracy. Since my

organization has continuously seen interest to investment in growth plans, identifying a new area of expansion is the hour of need with agility. (*HR Manager*) will be communicated on the final outcomes to ensure required potential resources are managed for the same. Revenue goals of organization: Customer churn is a major threat when coming to revenue goals of organization. Below data consolidation will provide a solid base that aids customer retention. 1. Data of current customer's contract years and its expiration 2. Customer satisfaction 3. Customer data on high profitable ones 4. Customer with multiple on-going project 5. Customer flexibility and co-operation 6. Goals and outcomes from Steering committee meetings 7. Team utilization and project commitment 8. Identify new areas of scope within customer. The data is ingested to know the customer's current scenario within the organization (*Strategy Analyst*). Strategies are dwelled to understand customer retention (*Data Engineers*). Come up with right strategy that prevents customer to build in-house provision of services my organization provides. Strategic decision based on evident outcomes from the above data to put into action is framed (*BI specialists and ML developers*). Being a leader will ensure to fine tune questionnaire's when and where required across all three business problems; ensure proper flow of information across different people in my organization happens with ease in an iterative way; test and training data will be formulated before each hypothesis is brought into action.

Modeling:

Choosing the right model is the key to derive most valuable insights out of the data reducing uncertainty. Addressing Business process strategy improvements: To understand loopholes in the current process of organization execution and derive 3 topmost business strategies that will focus in step one of improvements is the aim. The data collected has the true opinion from oneself that reflects real-time cross-department project handling status/issues. Since the data is more

descriptive and numeric, the ideal way is to understand where majority of data falls within segments. I go with supervised segmentation with tree- structured models with multiple attribute outcomes from the questionnaire for selecting best 3 attributes for business process improvements (target variable). Using the dataset and recursively applying attribute selection to find the best attributes for partitioning is done. Finding this model to be useful in visualizing segments to compare one against another for understanding correlations and relationships between each. From visualization part, it is better than regression fits that gets limited to two segments. As these are capable of producing probability estimation, I need to use the outcomes for predictive modelling in formulating new SOP's and practices that overcomes current scenario of working collaboratively in groups across departments. As a leader, I feel it's very crucial for my team to get the crux of the underlined problem statement. Meeting the business requirement is the key-hold required for the organization to focus which will in-turn increase the cross-departmental bonding. I will give enough spaces for the team to build their own analytical models and if they come up with something even better meeting the needs, I would be really happy to accept that as well. In the end, the results matter. Business area expansion: Aim is to come up with two applications which will aid in new business area expansion. The data from strategy analyst is used to find trend forecasting and to determine the strength of predictor variables. Here, I am using linear regression model to fit the data between scope of application's data collected versus its future trend in market to derive insights. My expectation with team is to finally come up with two applications which will be used across many customers in their daily business. Here, the selection criteria are more towards having a boundary margin line and understand which data falls above and below the line for finalizing the outcome. Here, there is no need to measure the impact of one over the other. Revenue goals of organization: Aim is to come up with strategies for existing

customer retention with the goal of discovering new scope of work in-turn providing value adds and increasing revenue of the organization. Making team members believe in the business value and its post effects which are crucial here. Once this is understood, clustering customers into categories like Simple, Medium, Complex will be easier. Formulating strategies and identifying clustering technique to fit the need is the key. Various customer-base specific policies for customer retention could be easily formed. Also, for simple and medium customers we can straight away dwell into identifying new scope of work eliminating complex customer at the moment.

Evaluation:

The predictive models that are in consideration for my business problems are

1. Supervised segmentation with Tree induction model;
2. Linear Regression;
3. Unsupervised method - Clustering

For the first business problem on identifying best attributes for business process improvements, I am proposing the Supervised segmentation using decision trees model. I believe the issues from the resultant questionnaire would be with having missing values in place and there will be chances of data overfitting. There are also few fields that have possibility of providing descriptive answers in place. Carefully pruning with this model with the intention of error reduction is the target. I am having the controls here on comparing the outcomes of information gain for entropy using base of log values with already available cost benefit matrix benchmarking expected benefit to the expected cost in-turn formulating the profits out for new business process strategy improvements. If this is not providing insights, I would re-meet the team of 20 persons considered for the representation in collecting the data; fix answers that are difficult for the classifiers in predicting the model. Once I measure the value of gain versus profit, I can provide with the evidence to the senior management for new strategic decisions in place. For the second business problem on

identifying the business area expansion, I am using linear regression model for forecasting and predicting new applications that justify investment in growth. In identifying the relationship, I am devising few conditional strategies like probability estimation to quantify the dependency on one over other. Devising interquartile range and identifying outliers with data collected are the most common challenges here. Here I am proposing performance measures on few key success factor's scores of the regression coefficient for individual applications and then comparing it with company's business area measurement goals and outcomes would be an ideal way. On the third business problem on increasing the revenue of the organization, I am implementing one of the unsupervised model approaches – Clustering using nearest neighbor classifier and K means algorithm. Using these techniques to pre-process the data by identifying similarity between customers will provide new mining opportunities. The exploratory technique which is challenging here is how to differentiate customer as simple, medium and complex for customer profiling. I foresee to identify expected value calculation where identifying false negative error holds much of the importance. I will plot a confusion matrix and derive distance matrix to identify similarity values for data in the questionnaire report and sort the threshold values for each of the targeted customer groups. The distances between datasets are usually correlated with centroids and dendrograms for the report. I can also use the already available data for training data sets and perform cross validation. Measuring the precision and accuracy from the model data will strategize my goals for revenue improvements.

Deployment:

Now that the project is reaching the final steps it is very important to know how to incorporate the models that fits data to improve the way the business yields value. As a first step in deployment process, I am coming up with deployment plan which has complete agenda with timelines and

ownership responsibilities on what they are going to do during go-live. For my business problem, I foresee monitoring it for measuring effectiveness throughout a period of time is the key. Say, if I am implementing a new business strategies for creating a new department in the organization or if I am in contact with customers formulating new proposals and contract extension, I feel this to be implemented with small group of people with whom we have better rapport and who are ready to come up with genuine feedbacks on how new changes and how they are bringing business value to them in BI perspective. If there are at all any issues and concerns projected, it is my solemn responsibility to ingest the feedback to the team in a positive note where everybody is on the scale of agility understanding iterations, rework and enhancing something better more than yesterday to the project is a good sign of personal and professional growth. There is also a greater need in updating process workflows and relevant documents which would act as a reference base. For new team formed or for new strategies in improving cross department bonding, it is extremely important to have SOP's updated with solid documentation base on new process changes. Having the lessons learned documented and incorporating them efficiently in future brings out a wider scope in understanding the user stories with much better clarity and focus. After a fixed timeframe, say 3 weeks, respective teams should sit together and clarify how things are going and what has improved since this implementation in place. Revamp of process changes, thought processes, etc... should be initiated and the clock needs to tick off from first always having the mindset of the activity as a collaborative team spirit work. Once these chunks of work are reviewed and revisited for multiple cycles; once the effectiveness measured has all user-story cases satisfied with successful live testing and feedbacks. Though the business problems remain the same, re-iterating the cycle using value driven approach based on feedbacks has lots of business strategic improvements. The reason is, earlier my team and I was primarily focusing on results driven

approaches based on values from the hypothesis and evaluation of models. Now re-doing with value-based approaches has better win-win situations. Having a mindset that focuses on driving the project for better values changed my overall perspective towards achieving my goal to a next better level. Finally, the values are reported in a meeting with management summarizing entire project with mention on team members efforts and dedication which has led to the complete success.

Conclusion:

New beginnings are often disguised as painful ending. But I strongly believe with meaningful data, successful insights are drawn solemnly. Analyzing employee feedback; understanding scope of lab application's current and future trend; visualizing customer's need will be strong pillars to lift my company in this competitive world. As a leader, it is effective when we give equal space for contribution of individuals inputs and suggestions. It is also effective when individuals in the team realize themselves as owner of the tasks which yields better productivity.

References:

- [1] Provost, F., & Fawcett, T. (2013). Data science for business: what you need to know about data mining and data-analytic thinking. Sebastopol: O'Reilly.
- [2] Collier, K. (2013). Agile analytics: a value-driven approach to business intelligence and data warehousing. Upper Saddle River, NJ: Addison-Wesley.
- [3] 13 Business Intelligence & Analytics Examples Illustrating the Value of BI. (2019, September 11). Retrieved from <https://www.datapine.com/blog/analytics-and-business-intelligence-examples/>.