THE ANALYTICS DEVELOPMENT PROCESS

Disclaimer: Analytics is not a purely linear process. It is expected that a practitioner will likely jump between different items outside of the order listed below.

1. Understand the business situation and core problem

*Describe the business situation*

*What is the core problem?*

*What is/are our target variable(s)? If none, what are we searching for?*

|  |  |  |
| --- | --- | --- |
| Supervised | Unsupervised | Mixed |

*List the deliverables, include relative weights on a 100-1 scale*

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

*What is the value proposition for this analytics project?*

*Link to and describe any external research*

1. Exploratory Data Analysis and Data Preparation

*Link to or copy in the data dictionary*

*What are the issues with data quality (not including outliers)?*

*Specifically, are there outliers in the data? How did we determine they were outliers? What will we do with them?*

*Describe any other initial cleaning and related assumptions*

*What are the issues with data integrity (including merging and hierarchies)?*

*Describe any data modifications made for integrity purposes and related assumptions*

*Describe any transformations made, including variable type conversion, creating dummy variables, binning continuous variables and mathematical function transformations. How may these help the analysis?*

*Is time a dimension in this data? If so, give your initial impression on any time series considerations, issues and solutions*

*For target variables (if any), comment on scatterplots and correlations for continuous variables, mean and variance analysis if one is dichotomous, and category characterizations for categorical variables. Are there non-linear relationships?*

*Provide other notes on any variables of interest (if there are targets, must discuss those) after reviewing summary tables and histograms and/or box plots. Link to a separate document if necessary*

*Describe any feature selection or reduction methods used*

*Will you construct any custom metrics or create any custom classifications?*

\*\*\*COPY SECTION 3 BEFORE WRITING AND PASTE BELOW 3\*\*\*

1. Modelling – Round 1

*What is/are the general purpose(s) of modeling for this round?*

*What type(s) of model(s) are being developed?*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Descriptive | Diagnostic | Predictive | Prescriptive | Time Series |

*Select the model(s) developed*

|  |  |  |  |
| --- | --- | --- | --- |
| UNSUPERVISED | | | |
| Association Rules | | | |
| Clustering | | | |
|  | K-means | Hierarchical | Gaussian Mixed Model |
| Principal Component Analysis | | | |
| Collaborative Filtering | | | |
| SUPERVISED | | | |
| Artificial Neural Networks | | | |
|  | Single Layer | Deep Learning | Convolutional/Boltzmann |
| Decision Tree / Random Forest | | | |
|  | Single Tree | Boosted Tree | Bagged Tree |
|  | Random Forest |  |  |
| K-Nearest-Neighbors | | | |
| *Classification* | | | |
| Logistic Regression | | | |
| Support Vector Machine | | | |
|  | Radial | Polynomial | Custom |
| Naïve Bayes | | | |
| Discriminant Analysis | | | |
| *Regression* | | | |
| Linear Regression | | | |
|  | Standard | Interaction | Non-Linear Independents |
|  | Moderation | Mediation | Hierarchical |
|  | Transformed | Non-Normal Errors |  |
| MARS – Multivariate Adaptive Regression Splines | | | |
| Penalized Regression | | | |
|  | Ridge | LASSO | Elastic Net |
| TIME SERIES | | | |
| Linear Regression | ARIMA | GARCH | Fourier Analysis |
| Artificial Neural Networks | Simple Moving Average | Seasonal Adjustment | Simple Exponential Smoothing |
| Holt’s Exponential Smoothing | Holt-Winters Exponential Smoothing | Survival Analysis |  |
| PRESCRIPTIVE | | | |
| Optimization | | | |
|  | Linear Programming | Integer Programming | Stochastic Programming |
|  | Combinatorial | Convex |  |
| Simulation | | | |

*List any other models developed*

|  |
| --- |
|  |
|  |
|  |
|  |
|  |

*Report any relevant diagnostics and/or model performance metrics*

*What new information did we learn about the variables?*

*What new information did we learn about the relationships?*

*Will any additional variables or data need to be brought into the analysis? If so, which variables and/or data?*

*Will any additional cleaning or transformations be required? If so, what cleaning and/or transformations?*

*Did the deliverables change? If so, what is new, modified or disregarded?*

*Is another modelling round necessary? Why?*

\*\*\*PASTE NEW MODEL ROUND HERE\*\*\*

1. Create Value, Provide Insights and Offer Recommendations

*List everything included in your deliverable package*

|  |
| --- |
|  |
|  |
|  |
|  |
|  |

*What insights did you derive from your analysis? If your end product is an automated algorithm, what specific functionality does it provide?*

*How are your insights actionable? What should a user do to put this information into practice? Are the actions operational or strategic? If your end product is an automated algorithm, how specifically can the algorithm be operationalized?*

*Otherwise, what value does your analysis/code provide to the user? How does it match the requested deliverables?*

*How is your analysis/code innovative or spark innovation in others?*

*Given more time, what could be done to expand upon this analysis/code?*