**Model Card - GLM**

## Model Details

* GLM - Generalized Linear Model
* Developed by John Nelder and R.W.M. Wedderbrun in 1972
* A flexible generalization of ordinary linear regression

### Form

* (Response Variable) is assumed to follow exponential family distribution with mean
* contains known covariates
* contains the coefficients to be estimated
* Fit by least squares and weighted least squares
  + Using SAS’s GLM procedure or R’s function

## Intended Use

GLM generalizes linear regression by allowing the linear model to be related to the response variable via a link function. No linear relationship is assumed between the response variable and explanatory variables.

* Binary Logistic Regression (Target is binary)
* Poisson Regression (For modelling events whose outcomes are counts)
* Advantages over traditional OLS regression
  + No need to transform the response to have a normal distribution
  + Choice of link function is separate from the choice of random component (more flexibility in modeling)
  + Models fitted via MLE, likelihood functions and parameter estimate benefit from asymptotic normal and chi-square distributions

## Factors

* Random Component
  + Specifies the probability distribution of the response variable
    - Normal Distribution in ordinary linear regression model
    - Binomial Distribution in binary logistic regression model
* Systematic Component
  + Specifies the explanatory variables () in the model
* Link Function or
  + Specifies the link between the random and the systematic components
  + Indicates how the expected value of the response variable relates to the linear combination of explanatory variables
  + for classical regression
  + for logistic regression

## Caveats and Recommendations

* The random component (response variable) doesn’t not have a separate error term
* Dependent variable does not need to be normally distributed, typically assumes a distribution from an exponential family (e.g. Binomial, Poisson, Multinomial, Normal, etc.)
* Assumes data are independently distributed
* Does not assume a linear relationship between the response variable and explanatory variables
* Assume a Linear relationship between the transformed expected response (in terms of the link function) and the explanatory variables (e.g. binary logistic regression )
* Errors are independent (not normally distributed)
* Parameter estimation with MLE