Table 1 - Regression metrics

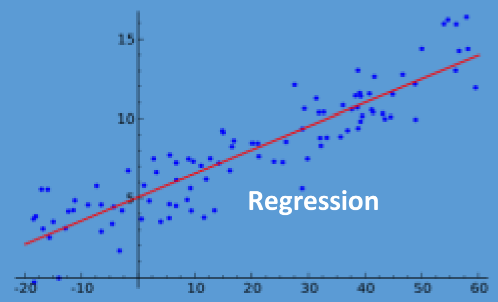
Group	Metric	Aplication	Example
 Regression	**MAE - Mean absolute erro**	Measure of difference between two continuous variables. MAE is the average vertical distance between each point and the identity line	25394.03
	MSE - Mean squared error	Measure the average of the squares of the errors—that is, the average squared difference between the estimated values and the actual value	1497295K
	RMSE - Root mean squared error	Frequently used measure of the differences between values (sample or population values) predicted by a model or an estimator and the values observed. The RMSD represents the square root of the second sample moment of the differences between predicted values and observed values or the quadratic mean of these differences	38694.91
	MAPE - Mean absolute percentage error	Measure of prediction accuracy of a forecasting method in statistics, for example in trend estimation, also used as a loss function for regression problems in machine learning. It usually expresses accuracy as a percentage	15.44
	RMSLE - Root Mean Squared Logarithmic Error	RMSLE is usually used when you don't want to penalize huge differences in the predicted and the actual values when both predicted and true values are huge numbers.	0.209787
	R-Squared - Coefficient of determination R^2	Is the proportion of the variance in the dependent variable that is predictable from the independent variable(s).	0.557863
	Adjusted R-Squared - adjust to features quantity	Adding new features to the model, the R-Squared value either increases or remains the same. R-Squared does not penalize for adding features that add no value to the model. So an improved version over the R-Squared is the adjusted R-Squared.	0.554811

Table 2.1 - Classification metrics

Group	Metric	Aplication	Example
True condition: Condition positive / Condition negative; Total population; Predicted condition: Predicted condition positive: True positive / False positive Type I error; Predicted condition negative: False negative Type II error / True negative	**TP - True Positive**	Hit	472
	TN - True Negative	Correct rejection	461
	FP - False positive (Type I error)	False alarm, Type I error	33
	FN - False negative	Miss, Type II error	34
Prevalence $= \frac{\Sigma \text{ Condition positive}}{\Sigma \text{ Total population}}$; Accuracy (ACC) $= \frac{\Sigma \text{ True positive} + \Sigma \text{ True negative}}{\Sigma \text{ Total population}}$	**PRV - Prevalence**	The prevalence measure helps us to measure the balance of the data within the total population	0.47
	ACC - Accuracy	It is the measure of success of the system, the total of positive and negative successes over the total population, it indicates what degree of success the model has	0.933
Positive predictive value (PPV), Precision $= \frac{\Sigma \text{ True positive}}{\Sigma \text{ Predicted condition positive}}$; False discovery rate (FDR) $= \frac{\Sigma \text{ False positive}}{\Sigma \text{ Predicted condition positive}}$; False omission rate (FOR) $= \frac{\Sigma \text{ False negative}}{\Sigma \text{ Predicted condition negative}}$; Negative predictive value (NPV) $= \frac{\Sigma \text{ True negative}}{\Sigma \text{ Predicted condition negative}}$	**PPV - Precision**	The PPV and NPV describe the performance of a diagnostic test or other statistical measure	0.9347
	NPV - Negative predictive value	The PPV and NPV describe the performance of a diagnostic test or other statistical measure	0.9313
	FDR - False Discovery rate (Type I)	The false discovery rate (FDR) is a method of conceptualizing the rate of type I errors in null hypothesis testing when conducting multiple comparisons	0.0653
	FOR - False omission rate		0.0687
True positive rate (TPR), Recall, Sensitivity, probability of detection, Power $= \frac{\Sigma \text{ True positive}}{\Sigma \text{ Condition positive}}$; False positive rate (FPR), Fall-out, probability of false alarm $= \frac{\Sigma \text{ False positive}}{\Sigma \text{ Condition negative}}$; False negative rate (FNR), Miss rate $= \frac{\Sigma \text{ False negative}}{\Sigma \text{ Condition positive}}$; Specificity (SPC), Selectivity, True negative rate (TNR) $= \frac{\Sigma \text{ True negative}}{\Sigma \text{ Condition negative}}$	**TPR - True positive rate - Recall - Sensitivity**	Sensitivity (also called the true positive rate, the recall, or probability of detection in some fields) measures the proportion of actual positives that are correctly identified as such.	0.9328
	FPR - False positive rate - Fall_out	false positive ratio (or false alarm ratio) is the probability of falsely rejecting the null hypothesis for a particular test	0.0668
	FNR - False negative rate	The false negative rate is the proportion of positives which yield negative test outcomes with the test, i.e., the conditional probability of a negative test result given that the condition being looked for is present.	0.0672
	TNR - True negative rate - Secificity	Specificity measures the proportion of actual negatives that are correctly identified as such	0.9332
Positive likelihood ratio (LR+) $= \frac{TPR}{FPR}$; Negative likelihood ratio (LR−) $= \frac{FNR}{TNR}$; Diagnostic odds ratio (DOR) $= \frac{LR+}{LR-}$; F_1 score $= 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$	**LR+ - Positive likelihood ratio**	Likelihood ratios are used for assessing the value of performing a diagnostic test. They use the sensitivity and specificity of the test to determine whether a test result usefully changes the probability that a condition (such as a disease state) exists.	13.9641
	LR- - Negative likelihood ratio		0.072
	DOR - Diagnostic odds ratio	Is a measure of the effectiveness of a diagnostic test.[1] It is defined as the ratio of the odds of the test being positive if the subject has a disease relative to the odds of the test being positive if the subject does not have the disease.	194

Developed by Kreilabs - www.kreilabs.com

Table 2.2 - Classification metrics

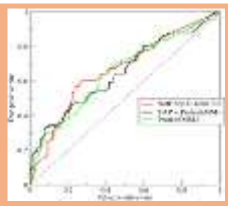
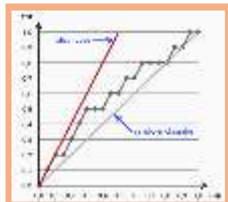
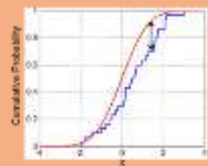
Group	Metric	Aplication	Example
$F_\beta = (1+\beta^2) \cdot \frac{\text{precision} \cdot \text{recall}}{(\beta^2 \cdot \text{precision}) + \text{recall}}$ The formula in terms of Type I and type II errors: $F_\beta = \frac{(1+\beta^2) \cdot \text{true positive}}{(1+\beta^2) \cdot \text{true positive} + \beta^2 \cdot \text{false negative} + \text{false positive}}$	**F1 score / Fbeta**	The F_1 score (also F-score or F-measure) is a measure of a test's accuracy. It considers both the precision p and the recall r of the test to compute the score: p is the number of correct positive results divided by the number of all positive results returned by the classifier, and r is the number of correct positive results divided by the number of all relevant samples (all samples that should have been identified as positive). The F_1 score is the harmonic mean of the precision and recall, where an F_1 score reaches its best value at 1 (perfect precision and recall) and worst at 0. The general formula for positive real β, where β is chosen such that recall is considered β times as important as precision. where β is non-negative number. When β=1, we are back to our familiar F1 score! Values of β much lower than 1, give more emphasis on precision, while values of β much higher than 1 give more emphasis on recall.	0.9337
Matthews correlation coefficient (MCC) $\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$ Informedness or Bookmaker Informedness (BM) $BM = TPR + TNR - 1$ Markedness (MK) $MK = PPV + NPV - 1$	**MCC - Mathews correlation coefficient**	Is used in machine learning as a measure of the quality of binary (two-class) classifications, it takes into account true and false positives and negatives and is generally regarded as a balanced measure which can be used even if the classes are of very different sizes	0.866
	BM - Bookmaker informedness	Informedness quantifies how informed a predictor is for the specified condition, and specifies the probability that a prediction is informed in relation to the condition (versus chance).	0.866
	MK - Markedness	Markedness quantifies how marked a condition is for the specified predictor, and specifies the probability that a condition is marked by the predictor (versus chance).	0.866
	**ROC curve analysis (Receiver Operating Characteristic)**	Is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. The ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings.	
	AUC (Area under curve)	When using normalized units, the area under the curve (often referred to as simply the AUC) is equal to the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one (assuming 'positive' ranks higher than 'negative').	0.937

Table 2.3 - Classification metrics and recomndations systems

Group	Metric	Aplication	Example
 	**Cummulative Gain chart**	Gain Charts are used for Evaluation of Binary Classifiers also it can be used for comparing two or more binary classifiers the chart shows tprtpr vs sup (predicted pos/total)	
	Lift chart	Lift charts show basically the same information as Gain charts. The lift is simply the ratio of the percentage of true positive to the percentage of true positive + negative.	
	Lift @decile chart	Similar lift chart, using real change values, not percentage	
	Kolomogorov Smirnov - KS	K-S or Kolmogorov-Smirnov chart measures performance of classification models.	
	Fowlkes–Mallows index	Geometric mean of precision and recall	
Precision and recall are classical evaluation metrics in binary classification algorithms and for document retrieval tasks. These metrics have been "Translated" to help us evaluate recommendation systems.	**recall@k**	In the context of recommendation systems we are most likely interested in recommending top-N items to the user. So it makes more sense to compute precision and recall metrics in the first N items instead of all the items. Recall at k is the proportion of relevant items found in the top-k recommendations	
	precision@k	Precision at k is the proportion of recommended items in the top-k set that are relevant	
	f1@k	F1 for k elements in recomendation systems	
	fbeta@k	Fbeta for k elements in recomendation systems	
	avgpre@k	Average precision for k element in recomendation systems	
$-\frac{1}{N}\sum_{i=1}^{N} y_i \cdot log(p(y_i)) + (1-y_i) \cdot log(1-p(y_i))$	**LogLoss**	Binary cross-entropy as loss function	
	Concordant/disconcordant ratio	Concordant ratio of more than 60% is considered to be a good model. It is primarily used to access the model's predictive power.	