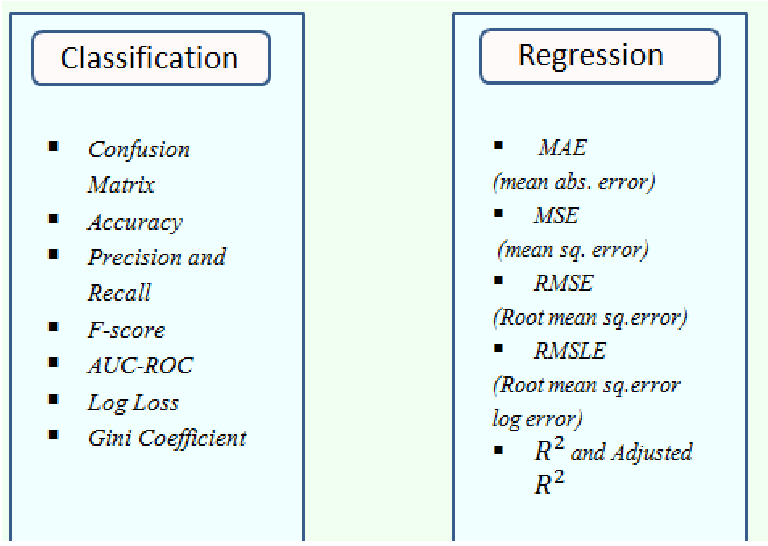


Performance metrices in ML

ML Evaluation Metrices

Monika Mittal
Accenture I Generative AI | Agentic AI | AI | NLP | LLM | 2X GCP | 1X Azure

March 14, 2024

Choosing the right performance metrics in machine learning is crucial for evaluating a model's effectiveness. Different metrics are suitable for various tasks, and understanding their strengths and weaknesses is essential. Here's a breakdown of commonly used metrics categorized by machine learning task type:

Classification Tasks:

- **Accuracy:** Measures the overall proportion of correctly classified instances.
- **Precision:** Indicates the ratio of true positive predictions to the total number of positive predictions (avoiding false positives).
- **Recall:** Represents the proportion of true positive predictions out of all actual positive cases (avoiding false negatives).
- **F1-Score:** A harmonic mean of precision and recall, providing a balanced view between the two.
- **Confusion Matrix:** Visualizes the model's performance, categorizing predictions into true positives, true negatives, false positives, and false negatives.

Regression Tasks:

- **Mean Squared Error (MSE):** Calculates the average squared difference between predicted and actual values. Lower MSE indicates better fit.
- **Root Mean Squared Error (RMSE):** Square root of MSE, interpretable in the same units as the data.
- **R-Squared (R^2):** Represents the proportion of variance in the dependent variable explained by the independent variable. A higher R^2 suggests a better fit.

Other Metrics:

- **Area Under the ROC Curve (AUC):** Measures the model's ability to distinguish between classes. Useful for imbalanced datasets.
- **Log Loss:** A metric used in classification with cost associated with misclassifications.

Choosing the Right Metric:

- **Task type:** Classification tasks use accuracy, precision, recall, F1-score, and confusion matrix. Regression tasks rely on MSE, RMSE, and R^2.
- **Cost of errors:** Some errors might be more critical than others. For example, in medical diagnosis, false negatives (missing a disease) could be much more detrimental than false positives.
- **Data balance:** In imbalanced datasets, accuracy might be misleading. AUC can be a better alternative.

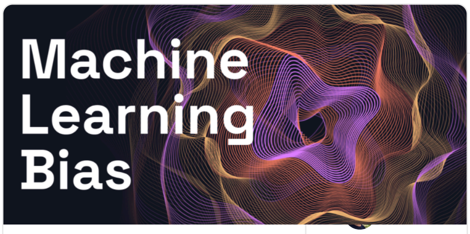
Additional Considerations:

- **Metrics can be complementary:** It's often beneficial to use multiple metrics to get a comprehensive picture of the model's performance.
- **Domain knowledge is crucial:** Understanding the real-world implications of errors helps prioritize relevant metrics.

By carefully selecting and interpreting performance metrics, data scientists can effectively evaluate their machine learning models and make informed decisions about their suitability for the intended task.

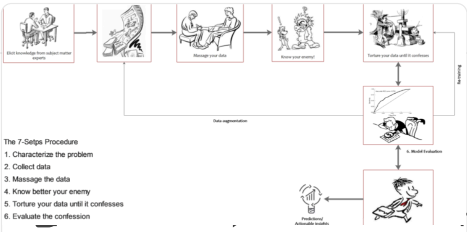
Monika Mittal
Accenture I Generative AI | Agentic AI | AI | NLP | LLM | 2X GCP | 1X Azure

More articles for you



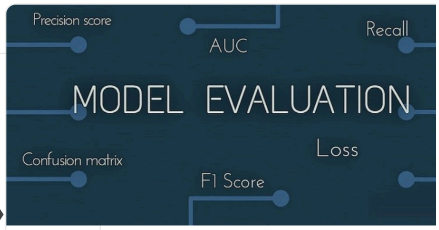
Understanding Bias in Feature Selection for Machine Learning Models
Rajesh Bairangula

3



The 7-Step Procedure of Machine Learning
Alaa Khamis, PhD

65 · 3 comments



Model Evaluation and Metrics: Understanding Different Evaluation Metrics for Model Performance...
Communication

9

Add a comment...