Propose a model monitoring pipeline and describe how you would track model drift in 500 words. Your answer can be saved as essay.pdf under the main repository.

Machine Learning (ML) Model Monitoring Pipeline: System Health, Data Quality, Model Quality and Relevance

Before tracking the model's quality, we must monitor the operationality and reliability of the model. This includes tracking system health issues with metrics such as input-output, memory and CPU usage to track model consumption, system uptime, disk utilisation, number of requests to ML API endpoint and latency per request.

Next, we monitor input data quality with percentage of missing values, type mismatches, duplicates, syntax, data type and format errors, data loss and more (Oladele, 2023).

We continuously assess model quality with performance metrics based on model type. For example, using accuracy, precision, and AUC-ROC for classification models, R-square, root-mean-squared-error and mean absolute error for regression models and normalised discounted cumulative gain, precision at K, mean average precision for ranking and recommendation models (EvidentlyAI, n.d.).

With the prediction outputs, we use stability metrics to capture the Prior Probability Shift, which is the distribution shift of dependent variables between training and production data or across various time frames of production data, such as Population Stability Index (PSI) and Divergence Index. We also monitor the Covariate Shift; distribution shift of independent variables, with metrics such as Characteristic Stability Index and Novelty Index (Saha, 2021), and lastly check for model drift.

To ensure that the model is non-discriminatory, we use bias and fairness metrics such as Predictive Parity to assess if the predictions are consistent across different groups, Equalized Odds to check if the error rates are comparable among different groups, Disparate Parity, Statistical Parity and other metrics that align with the context of the application.

Finally, we can define KPIs to measure the model's usefulness in meeting organisational goals. We will implement logging, report generation and data visualisation features to trigger alerts if any of the above metrics monitored fall below acceptable thresholds.

Model Drift: Concept Drift, Data Drift

To track concept drift, we can use Concept-adapting Very Fast Decision Tree (CVFDT) that adjusts tree structure by comparing accuracy of current and alternate subtrees to detect concept drift, replacing the subtree if necessary to adapt to changing data distributions (Jankowski et al., 2016). Another detection test would be the ADaptive WINdowing (ADWIN) algorithm. ADWIN divides the statistical window into two sub-windows and compares the average statistic to check if it corresponds to the same distribution. If the difference in averages is large, a drift is detected, triggering a replacement of the sub-window and initialization of a new one (Scikit, 2020).

As for data drift, we can use the Kolmogorov-Smirnov (KS) test, PSI and Z-Scores. The KS test compares cumulative distribution functions of two samples to determine if they come from the same underlying distribution (Jacobi, 2023). If the p-value is below a predetermined threshold, the two datasets have significantly different distributions and potential data drift. PSI measures differences between the expected and actual dataset distribution by comparing a set of predictor variables, a high PSI indicates larger differences and potential drift. Finally, the Z-Score detects changes in mean between training and live data, then compares this to the standard deviation of the variable in the training data. A large z-score indicates significant change in the mean and potential data drift (Gubkin, 2024).

Words: 497

Citations

Gubkin, A. (2024, March 6). ML model monitoring: Practical guide to boosting model performance. Aporia. https://www.aporia.com/learn/machine-learning-model/model-monitoring-101/

Jacobi, O. (2023, August 29). Model drift: What is it and ways to prevent it. Aporia. https://www.aporia.com/learn/data-drift/what-is-model-drift-and-5-ways-to-prevent-it/

Jankowski, D., Jackowski, K., & Cyganek, B. (2016). Learning decision trees from data streams with concept drift. Procedia Computer Science, 80, 1682–1691. https://doi.org/10.1016/j.procs.2016.05.508

Model monitoring for ML IN PRODUCTION: A comprehensive guide. Evidently AI - Open-Source ML Monitoring and Observability. (n.d.). https://www.evidentlyai.com/ml-in-production/model-monitoring#bias-and-fairness

Oladele, S. (2023, September 8). A comprehensive guide on how to monitor your models in production. neptune.ai. https://neptune.ai/blog/how-to-monitor-your-models-in-production-guide

Saha, P. (2021, September 16). MLOPS: Model monitoring 101. Medium. https://towardsdatascience.com/mlops-model-monitoring-101-46de6a578e03

Skmultiflow.drift_detection.adwin¶. scikit. (2020, June 17). https://scikit-multiflow.readthedocs.io/en/stable/api/generated/skmultiflow.drift_detection.ADWIN.html