

Contents

Model Cards a Data Sheets pro interní AI modely	1
1. Úvod do Model Cards a Data Sheets	1
1.1 Definice a účel	1
1.2 Povinnost dokumentace	1
2. Model Card Template	2
2.1 Základní informace	2
2.2 Training Information	2
2.3 Model Behavior a Limitations	3
2.4 Risk Assessment	3
3. Data Sheet Template	3
3.1 Dataset Composition	3
3.2 Collection Process	4
3.3 Data Content a Distribution	4
3.4 Privacy a Legal Compliance	5
3.5 Maintenance a Versioning	5
4. Governance a Approval Process	5
4.1 Review Process	5
4.2 Approval Workflow	6
4.3 Ongoing Monitoring	6
5. Template Checklist	6
5.1 Model Card Completeness	6
5.2 Data Sheet Completeness	6

Model Cards a Data Sheets pro interní AI modely

Verze: 1.0

Platnost od: [DATUM]

Příloha k: AI Usage Policy

Účel: Standardizace dokumentace pro interní AI modely

1. Úvod do Model Cards a Data Sheets

1.1 Definice a účel

Model Card je strukturovaný dokument popisující AI model včetně: - Zamýšleného použití a limitací - Performance metrik a testování - Etických úvah a bias analýzy - Training data a metodologie

Data Sheet dokumentuje dataset použitý pro training včetně: - Kompozice a zdroje dat - Collection metodologie - Preprocessing steps - Distribution a maintenance

1.2 Povinnost dokumentace

Kdy je nutná dokumentace: - Všechny interní ML/AI modely v production use - Modely použité pro business-critical rozhodování - Modely zpracovávající personal data - Experimentální modely s potential pro production

Výjimky: - Simple rule-based systémy bez ML komponenty - Third-party modely s existující dokumentací (vyžaduje link) - Proof-of-concept modely používané max. 30 dní

2. Model Card Template

2.1 Základní informace

Model Details:

Model Name: [Název modelu]
Version: [Verze, např. v1.2.3]
Date: [Datum vytvoření/update]
Model Type: [Classification/Regression/NLP/Computer Vision/etc.]
Architecture: [CNN/Transformer/Random Forest/etc.]
Framework: [TensorFlow/PyTorch/Scikit-learn/etc.]

Model Owner:

Primary Contact: [Jméno a email]
Team/Department: [Organizační jednotka]
Business Sponsor: [Business owner]
Technical Lead: [Tech lead kontakt]

Model Purpose:

Primary Use Case: [Hlavní účel použití]
Business Value: [Business benefit a ROI estimation]
Deployment Environment: [Production/Staging/Development]
Integration Points: [Systémy používající model]

2.2 Training Information

Training Data:

Dataset Name: [Název dataset]
Data Sources: [Zdroje dat - internal/external/public]
Data Collection Period: [Od-do datum]
Data Size: [Počet records, velikost]
Data Quality Score: [Kvalita dat 1-10 s vysvětlením]

Training Configuration:

Training Period: [Doba training]
Compute Resources: [GPU/CPU specs, cloud provider]
Training Duration: [Celkový čas training]
Hyperparameters: [Klíčové hyperparameters]
Cross-validation: [Typ validation použitý]

Performance Metrics:

Primary Metric: [Hlavní metrika - accuracy/F1/AUC/etc.]
Training Performance: [Hodnoty na training set]
Validation Performance: [Hodnoty na validation set]

Test Performance: [Hodnoty na test set]
Benchmark Comparison: [Srovnání s baseline/industry standard]

2.3 Model Behavior a Limitations

Intended Use:

Primary Applications: [Seznam hlavních use cases]
Target Users: [End users, internal teams]
Input Requirements: [Formát a requirements input dat]
Output Interpretation: [Jak interpretovat output]
Decision Thresholds: [Confidence thresholds pro decision making]

Known Limitations:

Data Limitations: [Omezení training dat]
Performance Degradation: [Znamé situace kde model selhává]
Scope Limitations: [Co model NEMŮŽE dělat]
Temporal Limitations: [Časová omezení, concept drift]
Technical Constraints: [Hardware/software requirements]

Bias a Fairness Analysis:

Protected Attributes: [Demografické/chráněné skupiny analyzované]
Bias Testing Results: [Výsledky bias testing]
Fairness Metrics: [Demographic parity/Equal opportunity/etc.]
Mitigation Strategies: [Kroky pro reduction of bias]
Ongoing Monitoring: [Monitoring plan pro bias detection]

2.4 Risk Assessment

Risk Categories:

High Risk Factors: [Faktory s high impact]
Medium Risk Factors: [Moderate impact faktory]
Low Risk Factors: [Minor impact faktory]
Mitigation Measures: [Konkrétní risk mitigation steps]

Ethical Considerations:

Potential Harm: [Možné negative dopady]
Stakeholder Impact: [Dopad na různé stakeholder skupiny]
Privacy Implications: [Privacy risks a safeguards]
Transparency Requirements: [Explainability requirements]

3. Data Sheet Template

3.1 Dataset Composition

Basic Information:

Dataset Name: [Název dataset]

Version: [Verze dataset]
Creation Date: [Datum vytvoření]
Size: [Počet instances/samples]
Features: [Počet a typ features]

Data Sources:

Primary Sources: [Hlavní zdroje dat]
Secondary Sources: [Dodatečné zdroje]
Data Collection Method: [Automated/Manual/Survey/etc.]
Geographic Coverage: [Geografické pokrytí]
Temporal Coverage: [Časové období pokryté daty]

Instance Information:

Instance Definition: [Co představuje jeden instance]
Total Instances: [Celkový počet]
Missing Data: [Procento a typ missing values]
Duplicate Detection: [Metodologie pro duplicate detection]
Quality Metrics: [Data quality assessment results]

3.2 Collection Process

Data Collection:

Collection Timeframe: [Období sběru dat]
Collection Methodology: [Detailní popis metodologie]
Sampling Strategy: [Random/Stratified/Convenience/etc.]
Data Collectors: [Kdo data sbíral - internal/external]
Collection Tools: [Software/hardware použité]

Data Preprocessing:

Cleaning Steps: [Kroky data cleaning]
Transformation Steps: [Normalization/encoding/etc.]
Feature Engineering: [Nové features vytvořené]
Filtering Criteria: [Kritéria pro exclusion dat]
Quality Checks: [Validation steps provedené]

3.3 Data Content a Distribution

Statistical Summary:

Descriptive Statistics: [Mean/median/mode pro numeric features]
Distribution Analysis: [Skewness/outliers/etc.]
Class Distribution: [Balance/imbalance v target variable]
Correlation Analysis: [Významné korelace mezi features]

Data Categories:

Categorical Features: [Seznam s unique values count]
Numerical Features: [Range a distribution info]
Text Features: [Language/length statistics]

Date/Time Features: [Temporal distribution]

3.4 Privacy a Legal Compliance

Data Privacy:

Personal Data: [Obsahuje/neobsahuje personal data]

Sensitive Categories: [GDPR special categories]

Anonymization: [Metody anonymization použité]

Consent Basis: [Legal basis pro processing]

Data Subject Rights: [Jak jsou zajištěna práva data subjects]

Legal Compliance:

GDPR Compliance: [Status a dokumentace]

Data Protection Impact Assessment: [DPIA provedena/requirements]

Cross-border Transfers: [International transfers pokud applicable]

Retention Policy: [Data retention timeline]

Deletion Procedures: [Right to erasure implementation]

3.5 Maintenance a Versioning

Data Updates:

Update Frequency: [Jak často jsou data aktualizována]

Update Process: [Proces pro data updates]

Version Control: [Versioning strategy]

Change Documentation: [Jak jsou změny dokumentovány]

Backward Compatibility: [Impact changes na existing models]

Data Lineage:

Source Traceability: [Sledovatelnost až ke zdrojům]

Processing History: [Historie všech transformací]

Dependencies: [Závislosti na jiných datasetech]

Impact Analysis: [Analýza dopadu změn]

4. Governance a Approval Process

4.1 Review Process

Technical Review:

Code Review: [Peer review training kódu]

Architecture Review: [Review model architektury]

Performance Review: [Validace performance claims]

Security Review: [Security assessment modelu]

Business Review:

Use Case Validation: [Validace business use case]

ROI Assessment: [Return on investment analysis]

Risk Assessment: [Business risk evaluation]
Stakeholder Approval: [Sign-off od key stakeholders]

4.2 Approval Workflow

Approval Stages: 1. **Technical Approval:** Data Science team lead 2. **Security Approval:** InfoSec team 3. **Privacy Approval:** DPO (pokud applicable) 4. **Business Approval:** Business sponsor 5. **Final Approval:** CTO/IT Director

Documentation Requirements:

Model Card: [Kompletní a approved]
Data Sheet: [Kompletní a approved]
Risk Assessment: [Completed risk analysis]
Test Results: [Validation a performance testing]
Deployment Plan: [Production deployment strategy]

4.3 Ongoing Monitoring

Performance Monitoring:

Production Metrics: [Monitoring v production]
Model Drift Detection: [Data a concept drift monitoring]
Performance Degradation: [Alerting na performance issues]
Feedback Loop: [Incorporation of user feedback]

Compliance Monitoring:

Bias Monitoring: [Ongoing bias detection]
Privacy Compliance: [Regular privacy audits]
Data Quality: [Ongoing data quality monitoring]
Documentation Updates: [Regular documentation reviews]

5. Template Checklist

5.1 Model Card Completeness

Povinné sekce: - ☐ Model Details a Owner information - ☐ Training Data a Configuration - ☐ Performance Metrics - ☐ Intended Use a Limitations - ☐ Bias a Fairness Analysis - ☐ Risk Assessment - ☐ Ethical Considerations

Doporučené doplnění: - ☐ Model Architecture diagram - ☐ Performance comparison chart - ☐ Error analysis - ☐ User guidelines - ☐ Maintenance schedule

5.2 Data Sheet Completeness

Povinné sekce: - ☐ Dataset Composition - ☐ Collection Process - ☐ Data Content a Distribution - ☐ Privacy a Legal Compliance - ☐ Maintenance a Versioning

Doporučené doplnění: - ☐ Statistical visualizations - ☐ Data sample examples (anonymized) - ☐ Collection timeline - ☐ Data lineage diagram - ☐ Quality metrics dashboard

Template připraven: RRSof AI Technologies

Pro použití: Enterprise AI governance

Další podpora: AI Ethics a Model Governance týmy

Tento template je založen na industry best practices (Google Model Cards, Microsoft Datasheets for Datasets). Přizpůsobte podle specifických potřeb vaší organizace.