A Mashup-Based Approach for Predictive Compliance Monitoring*

Adrián Romero-Flores[1,*], Alfonso E. Márquez-Chamorro[1,2] and Cristina Cabanillas[1,2]

[1]Smart Computer Systems Research and Engineering Lab (SCORE), Universidad de Sevilla, Spain
[2]I3US Institute, Universidad de Sevilla, Spain

Abstract

Predictive Compliance Monitoring (PCM) is an emerging field that integrates predictive techniques with compliance monitoring to ensure business processes adhere to regulatory and organizational standards. Existing research in this area has been limited, as current approaches mainly focus on monitoring Service Level Agreements (SLAs) or restrict predictions to remaining execution time. This paper introduces a framework for PCM that utilizes multiple predictive process monitoring (PPM) models on compliance mashups. The framework forecasts key process indicators, including next event predictions, remaining execution time, and process outcomes, while simultaneously evaluating compliance with predefined rules. This approach advances automated compliance monitoring and highlights key challenges and future research opportunities in the PCM field.

Keywords

predictive compliance monitoring, predictive process monitoring, compliance monitoring, business process management, compliance mashups, compliance checking

1. Introduction

Ensuring compliance of business processes with internal policies and external regulations – such as GDPR for data protection, HIPAA for healthcare data privacy, or ISO 27001 for information security management – is a critical challenge for organizations. Compliance violations can lead to legal penalties, financial losses and operational inefficiencies. Traditional compliance monitoring techniques typically focus on detecting non-compliant behavior either retrospectively (*post-mortem*) or during process design and execution (*pre-mortem*) [1].

However, these reactive approaches are often insufficient in mitigating compliance risks before violations occur. A key challenge in compliance monitoring is integrating predictive capabilities that anticipate potential violations in advance. Developing an approach that can generalize across different compliance requirements and provide early detection while ensuring accuracy, interpretability, and scalability remains an open research problem.

Predictive Process Monitoring (PPM) offers a promising approach to addressing this challenge. PPM leverages historical data from process executions stored in event logs to forecast the future behavior of running process instances [2]. These predictions can take different forms, such as estimating remaining execution time, forecasting outcomes, or anticipating the next process event. These predictions act as key performance indicators for ongoing processes, offering valuable insights to businesses.

In this idea paper, we introduce a conceptual framework for Predictive Compliance Monitoring (PCM) that leverages *compliance mashups* [3] to generate partial PPM predictions. Our approach

In: Janis Grabis, Yves Wautelet, Emanuele Laurenzi , Hans-Friedrich Witschel, Peter Haase, Marco Montali, Cristina Cabanillas, Andrea Marrella, Manuel Resinas, Karolin Winter. Joint Proceedings of HybridAIMS and CAI Workshops. Co-located with CaiSE 2025.

*This publication is part of the R&D projects PID2021-126227NB-C21 (PERSEO) and PID2022-140221NB-I00 (TAPIOCA) funded by MICIU/AEI/10.13039/501100011033/FEDER/UE; and TED2021-131023B-C22 (ORCHID) and PDC2022-133521-I00 (STATUS) funded by MICIU/AEI/10.13039/501100011033/European Union NextGenerationEU/PRTR.

*Corresponding author.
aromero17@us.es (A. Romero-Flores); amarquez6@us.es (A. E. Márquez-Chamorro); cristinacabanillas@us.es (C. Cabanillas)
0009-0009-3755-1731 (A. Romero-Flores); 0000-0002-8243-0404 (A. E. Márquez-Chamorro); 0000-0001-9182-8847 (C. Cabanillas)

tackles key challenges in the PCM field, including the generalization of compliance requirements beyond SLA-specific constraints and the integration of multiple predictive models to handle more complex compliance rules. Furthermore, we identify key research challenges in the PCM field, offering a foundation for future research.

The remaining of the paper is structured as follows. Section 2 outlines related work and the limitations of existing approaches that motivate this work. Section 3 presents our proposed framework for PCM based on compliance mashups. Finally, Section 4 reflects on the approach, its scope and limitations, the challenges ahead and plans for future work in this research area.

2. Motivation

Predictive compliance monitoring is a largely unexplored research area that focuses on forecasting potential violations of compliance constraints in business processes. A recent survey on PCM [4] highlights the limited research in this domain and identifies key studies contributing to the field. As discussed in [4], two primary approaches to PCM exist: (i) Predicate Prediction, which predicts whether a constraint is fulfilled, and (ii) PCM based on PPM, which builds a compliance monitoring system on top of PPM results. For example, Predicate Prediction might be used to determine whether an invoice approval will meet a regulatory deadline before the process is completed. In contrast, a PCM approach on PPM would first predict the expected remaining time of an approval process and then assess whether this prediction aligns with compliance constraints. The key distinction lies in Predicate Prediction providing a direct compliance verdict, whereas PCM based on PPM relies on intermediate process forecasts to infer compliance status.

Rinderle-Ma et al. [4] identified seven key studies that can be classified as PCM. Leitner et al. [5, 6] predict Service Level Agreements (SLA) violations using remaining time prediction, qualifying as a PPM approach. In [6], they also introduce preventive methods to ensure that predictions adhere to the defined Service Level Objectives (SLOs). Khan et al. [7] provide a generic model for compliance prediction in business processes, utilizing trace activity and time data for binary classification. Cicotti et al. [8] propose a Quality of Service (QoS) predictor for SLA compliance monitoring based on a probabilistic model. Rodríguez et al. [9] employ decision trees to explain compliance root causes and use them to predict whether a process is compliant. Comuzzi et al. [10] present a metric for SLA predictive monitoring by forecasting boolean SLOs. Ivanović et al. [11] focus on predicting QoS at runtime to support service orchestration.

These approaches suffer from several limitations that prevent them from functioning as a comprehensive PCM system. Specifically, existing methods typically fall into one of three categories: (i) they focus solely on remaining time prediction in PPM, (ii) they are restricted to SLA compliance, or (iii) they operate only at the level of individual process instances [4]. As a result, these approaches rely on relatively simple compliance rules, lacking the ability to handle more complex compliance constraints that involve multiple features or events.

This gap underscores the need for further research to develop PCM solutions that move beyond SLA-specific compliance monitoring, incorporate advanced predictive capabilities, and support the enforcement of complex compliance rules spanning multiple process features or events.

3. Our Proposal

We propose a conceptual framework for PCM that leverages *compliance mashups* [3] for defining and automatically monitoring compliance rules. Our approach generates PPM models on these mashups, enabling partial predictions of the rules and ultimate unifying the obtained results to verify compliance.

A mashup is a data-driven workflow (a.k.a. *data flow*) built with information from one or more data sources that might be transformed and propagated to produce a desired output in a reusable User Interface (UI) [12]. Compliance mashups [3] refer to mashups specifically designed for compliance checking purposes. These mashups facilitate the automated verification of compliance rules at design

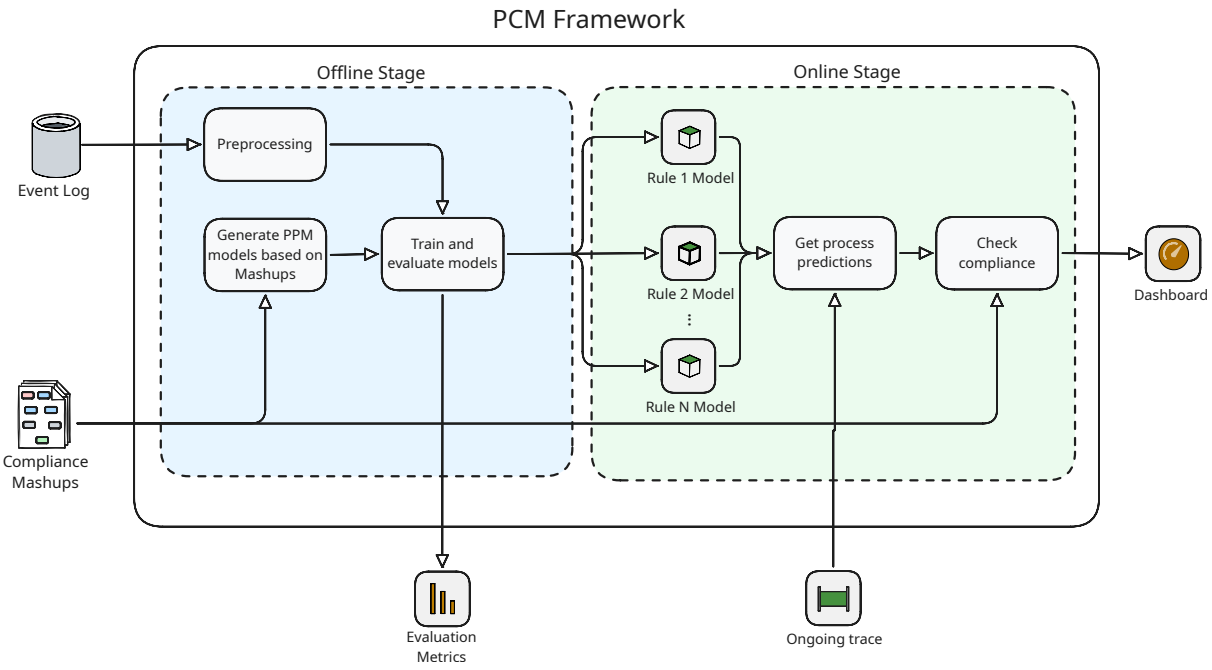

Figure 1: Architecture Diagram of the Predictive Compliance Monitoring proposed.

time or run time by integrating heterogeneous data sources, applying necessary transformations, and producing compliance evaluation results in an interpretable format. Each mashup operationalizes one compliance rule, although mashups can be embedded into domain-specific components to adapt to different granularities and enhance scalability. Tools like *STATUS* [13] leverage the mashup-based compliance management framework presented in [3] to build a low-code compliance monitoring system.

Figure 1 depicts the architecture of our proposed framework for PCM. We assume that all compliance rules can be verified based on data from event logs. This assumption is generally adopted by existing compliance monitoring approaches [14, 15]. Our framework integrates PPM, a well established discipline that applies machine learning techniques to forecast process behavior or outcomes. Specifically, we consider predictive models generated by machine learning techniques such as neural networks or decision trees, which estimates the next process event(s), next event time, remaining execution time, process outcome, or any combination of these [2].

The framework comprises two stages. For a compliance rule, the offline stage generates PPM models based on historical event logs containing information about the respective process executions, and on the compliance mashup that monitors the rule. First, the system parses the mashup to identify the components of the compliance rule. Subsets of the mashup components can be used to infer intermediate PPM models. These intermediate models serve as building blocks for the final compliance prediction. After constructing the individual models, the system preprocesses the event log, trains each rule-specific model, and evaluates its performance using well-established evaluation metrics. We consider classification metrics such as accuracy, precision, recall, and F1-score for categorical predictions (e.g., predicting the next activity). For numerical predictions, such as remaining execution time, we use regression metrics like Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).

In the online stage, the trained models generate compliance predictions for ongoing process instances. The system processes the latest instance data through the corresponding predictive models, producing probabilistic forecasts of upcoming events, execution times, and process outcomes. These predictions are then integrated into compliance mashups, enabling real-time assessment of whether the process is likely to remain compliant or violate predefined rules.

We exemplify the use of the framework with a simple scenario. Consider an industrial production plant where maintaining product quality and preventing overheating are critical. The plant follows a structured cooling and quality assurance process to ensure that materials are processed within

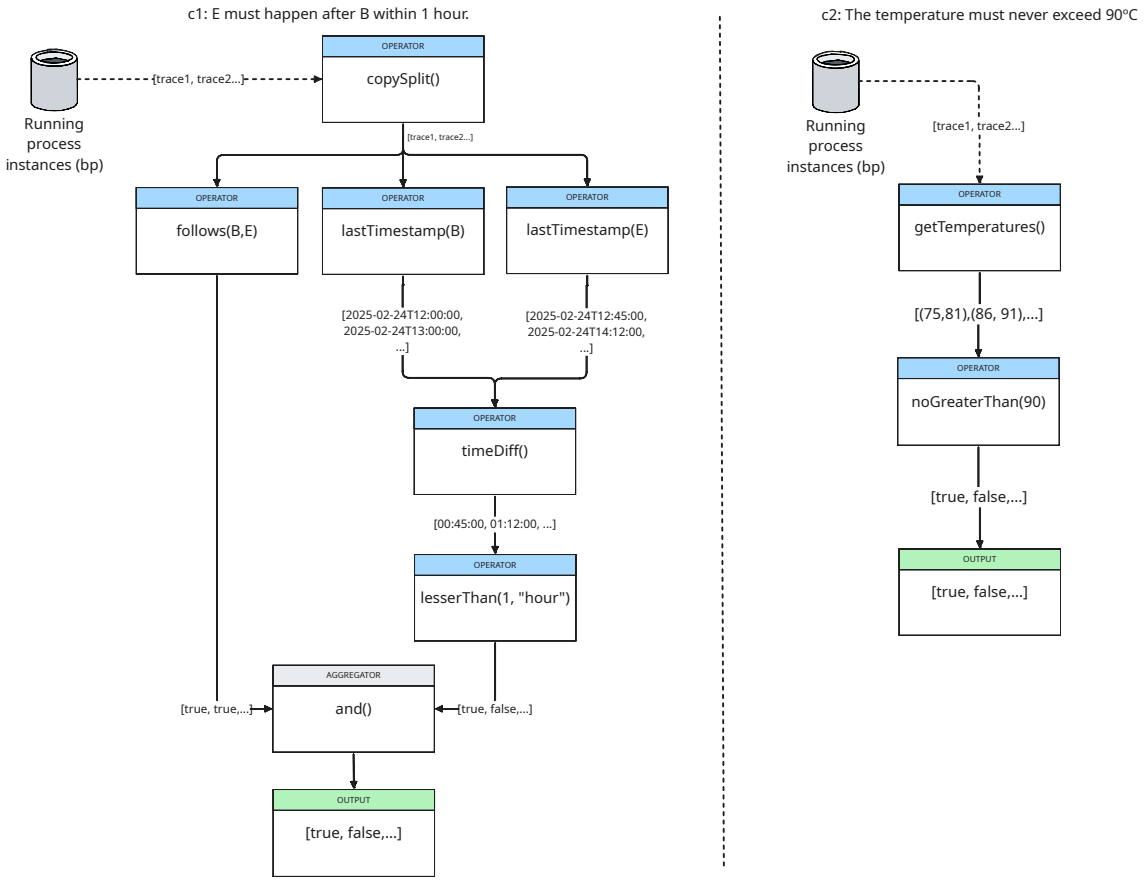

Figure 2: Compliance mashups for rules *c1* and *c2*.

predefined temperature and time regulations. Each batch of materials undergoes a series of steps, starting with (A) material preparation, followed by (B) heat treatment, (C) cooling, (D) quality inspection, and finally (E) packaging. Throughout the process, compliance with two key rules must be ensured. First (*c1*), once a batch enters the heat treatment phase, packaging must be completed within 60 minutes to prevent material degradation. Second (*c2*), the material's temperature must remain below 90°C to avoid safety hazards.

The process event log entries include a unique identifier for each process instance that enable the recognition of process traces, labeled *caseID*; the specific process activity to which the event refers, named *activity*; the date and time when the activity occurred, marked as *timestamp*; and a numerical value for temperature measurements, identified as *temperature*.

The compliance mashups for rules *c1* and *c2* are depicted in Figure 2. These mashups consist of nodes that retrieve and transform information from process traces, enabling the generation and training of the necessary PPM models based on historical data.

The predictive models required for our example are as follows. For rule *c1*, the system includes two predictors. The *Next activity predictor* forecasts the subsequent process activities on the `follows(B,E)` node in mashup *c1*, which detects if the activity B is followed eventually by the activity E. The *Next event time predictor* estimates the time difference between events, utilizing the `lastTimestamp(B)`, `lastTimestamp(E)`, and `timeDiff()` nodes. For Rule *c2*, the *Outcome predictor* is used, which infers the process outcome from the `getTemperatures()` node in mashup *c2*.

When a process instance is actively running, the system employs the previously trained models to generate compliance predictions in real time. Based on the models associated with each compliance mashup, the system evaluates the likelihood of future compliance or violation. Figure 3 illustrates an example of the online stage as seen in Figure 1, where a running trace undergoes predictive compliance assessment. *Rule c1 Model* forecasts the next activity and expected event timestamp for the monitoring

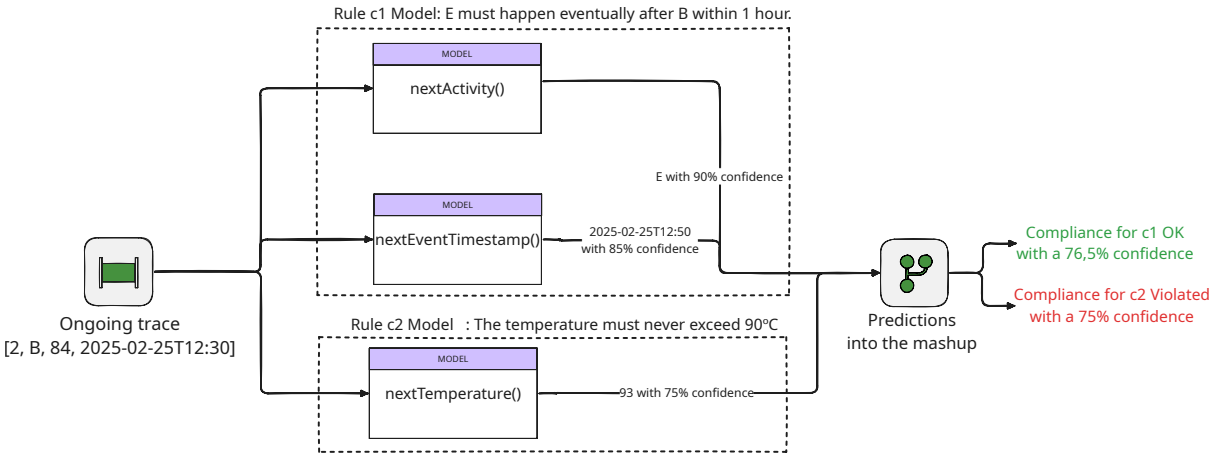

Figure 3: Predictions over a running trace.

of *c1*; and *Rule c2 Model* forecasts the temperature for the monitoring of *c2*.

These results are then integrated into the compliance mashup to determine the compliance degree for each rule. The compliance degree is a quantitative measure of how closely the predicted process execution aligns with the compliance rules. The compliance degree can be expressed as a binary classification (compliant / non-compliant) with a probability score reflecting the likelihood of the prediction. For example, in Figure 3 the compliance degree for rule *c1* is classified as compliant with a probability of 76.5%. This combined probability for *c1* is derived from the confidences of the constituent predictions; in this instance, it corresponds to the product of the 90% confidence for the predicted next activity (E) and the 85% confidence for the predicted timestamp ($0.90 \cdot 0.85 = 0.765$). For rule *c2*, it is categorized as non-compliant with a probability of 75%.

4. Conclusions

This idea paper proposes a novel approach to PCM using compliance mashups. By integrating predictive models with compliance mashups, this approach enhances the coverage of the Compliance Monitoring Functionalities (CMFs) related to PCM outlined in [4]. Specifically, it supports previously uncovered functionalities such as non-atomic activities (CMF4), life cycle considerations (CMF5), multiple instance constraints (CMF6), root cause analysis (CMF9) and compliance degree calculation (CMF10).

The main limitations of this approach are, first, the lack of integration with information systems that do not rely on event logs, which restricts the range of applicable data sources. Second, the types of predictions supported are inherently constrained by the capabilities of PPM, limiting the scope of compliance violations that can be anticipated.

This work also unveils several challenges that remain open. One of the key issues is implementing efficient online (re-)training mechanisms for each predictive model to adapt to evolving processes dynamically. The second challenge lies in determining the compliance degree, as common regression models do not inherently provide a confidence measure. Some methods, such as those proposed in [16, 17], offer estimation techniques for this measure. The third open challenge is defining the number of future events to predict. This introduces the need for sophisticated predictive models, such as Long Short-Term Memory (LSTM) networks, which can generate sequences of predictions rather than isolated forecasts. Lastly, the challenge of inferring the appropriate predictive models from the mashups emerges. This involves determining which mashup components are necessary for compliance predictions and how to efficiently combine them.

Future research should focus on addressing these limitations and challenges, and developing scalable architectures for real-world applications. By tacking these issues, compliance monitoring systems can transition from reactive approaches to fully proactive PCM systems, enabling early compliance insights,

reducing operational risks, and ensuring greater regulatory adherence.

Declaration on Generative AI

During the preparation of this work, the author(s) used ChatGPT in order to: Grammar and spelling check, Paraphrase and reword. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] L. T. Ly, F. M. Maggi, M. Montali, S. Rinderle-Ma, W. M. van der Aalst, Compliance monitoring in business processes: Functionalities, application, and tool-support, Information Systems 54 (2015) 209–234. doi:https://doi.org/10.1016/j.is.2015.02.007.

[2] A. E. Márquez-Chamorro, M. Resinas, A. Ruiz-Cortés, Predictive monitoring of business processes: A survey, IEEE Transactions on Services Computing 11 (2018) 962–977. doi:10.1109/TSC.2017.2772256.

[3] C. Cabanillas, M. Resinas, A. Ruiz-Cortés, A Mashup-Based Framework for Business Process Compliance Checking, IEEE Transactions on Services Computing 15 (2022) 1564–1577. doi:10.1109/TSC.2020.3001292.

[4] S. Rinderle-Ma, K. Winter, J.-V. Benzin, Predictive compliance monitoring in process-aware information systems: State of the art, functionalities, research directions, Information Systems 115 (2023) 102210. URL: https://www.sciencedirect.com/science/article/pii/S0306437923000467. doi:https://doi.org/10.1016/j.is.2023.102210.

[5] P. Leitner, J. Ferner, W. Hummer, S. Dustdar, Data-driven and automated prediction of service level agreement violations in service compositions, Distributed and Parallel Databases 31 (2013) 447–470. doi:10.1007/s10619-013-7125-7.

[6] P. Leitner, A. Michlmayr, F. Rosenberg, S. Dustdar, Monitoring, Prediction and Prevention of SLA Violations in Composite Services, in: IEEE International Conference on Web Services (ICWS), 2010, pp. 369–376. doi:10.1109/ICWS.2010.21.

[7] N. Khan, Z. Ali, A. Ali, S. McClean, D. Charles, P. Taylor, D. Nauck, A Generic Model for End State Prediction of Business Processes Towards Target Compliance, 2019, pp. 325–335. doi:10.1007/978-3-030-34885-4_25.

[8] G. Cicotti, L. Coppolino, S. D'Antonio, L. Romano, Runtime Model Checking for SLA Compliance Monitoring and QoS Prediction, J. Wirel. Mob. Networks Ubiquitous Comput. Dependable Appl. 6 (2015) 4–20. doi:http://dx.doi.org/10.22667/JOWUA.2015.06.31.004.

[9] C. Rodríguez, P. Silveira, F. Daniel, F. Casati, Analyzing Compliance of Service-Based Business Processes for Root-Cause Analysis and Prediction, 2010, pp. 277–288. doi:10.1007/978-3-642-16985-4_25.

[10] M. Comuzzi, A. E. Marquez-Chamorro, M. Resinas, A hybrid reliability metric for sla predictive monitoring, in: ACM/SIGAPP Symposium on Applied Computing, 2019, p. 32–39. URL: https://doi.org/10.1145/3297280.3297285. doi:10.1145/3297280.3297285.

[11] D. Ivanović, M. Carro, M. Hermenegildo, Constraint-Based runtime prediction of SLA violations in service orchestrations, in: International Conference on Service-Oriented Computing (ICSOC), 2011, p. 62–76. doi:10.1007/978-3-642-25535-9_5.

[12] F. Daniel, M. Matera, Mashups: Concepts, Models and Architectures, Springer, 2014.

[13] Á. Bernal, F. Montero, C. Cabanillas, P. Fernandez, M. Resinas, STATUS: A Low-Code Business Process Compliance Management System, in: BPM Demos, volume 3758, 2024, pp. 141–145. URL: https://ceur-ws.org/Vol-3758/paper-26.pdf.

[14] A. Awad, E. Pascalau, M. Weske, Towards Instant Monitoring of Business Process Compliance, EMISA Forum 30 (2010) 10–24.

[15] D. Knuplesch, M. Reichert, A. Kumar, A framework for visually monitoring business process compliance, Inf. Syst. 64 (2017) 381–409. doi:10.1016/j.is.2016.10.006.

[16] Z. Bosnić, I. Kononenko, Comparison of approaches for estimating reliability of individual regression predictions, Data Knowledge Engineering 67 (2008) 504–516. doi:10.1016/j.datak.2008.08.001.

[17] Z. Bosnić, I. Kononenko, Estimation of individual prediction reliability using the local sensitivity analysis, Applied Intelligence 29 (2008) 187–203. doi:10.1007/s10489-007-0084-9.