Nirdizati: A Web-Based Tool for Predictive Process Monitoring

Kerwin Jorbina[1], Andrii Rozumnyi[1], Ilya Verenich[1,2], Chiara Di Francescomarino[3], Marlon Dumas[1], Chiara Ghidini[3], Fabrizio Maria Maggi[1], Marcello La Rosa[3], and Simon Raboczi[3]

[1] University of Tartu, Estonia
{kerwin.jorbina,andriiro,marlon.dumas,f.m.maggi}@ut.ee
[2] Queensland University of Technology, Australia
{ilya.verenich,m.larosa,simon.raboczi}@qut.edu.au
[3] FBK IRST, Trento, Italy
{dfmchiara,ghidini}@fbk.eu

Abstract. This paper introduces Nirdizati: A web-based application for generating predictions about running cases of a business process. Nirdizati is a configurable full-stack web application that supports users in selecting and tuning prediction methods from a list of implemented algorithms and enables the continuous prediction of various performance indicators at runtime. The tool can be used to predict the outcome, the next events, the remaining time, or the overall workload per day of each case of a process. For example, in a lead-to-order process, Nirdizati can predict which customer leads will convert to purchase orders and when. In a claim handling process, it can predict if a claim decision will be made on time or late. The predictions, as well as real-time summary statistics about the process executions, are presented in a dashboard that offers multiple visualization options. Based on these predictions, process participants can act proactively to resolve or mitigate potential process performance violations. The target audience of this demonstration includes process mining researchers as well as practitioners interested in exploring the potential of predictive process monitoring.

Keywords: Process Mining, Predictive Process Monitoring, Machine Learning

1 Introduction

Predictive Process Monitoring [1] is an emerging paradigm based on the continuous generation of predictions about the future values of user-specified performance indicators of a currently running process execution. In this paradigm, a user defines the type of predictions they are interested in and provides a set of historical execution traces. Based on the analysis of these traces, the idea of predictive monitoring is to continuously provide the user with predictions and estimated values of the performance indicators. Such predictions generally depend both on: (i) the sequence of activities executed in a given case; and (ii) the values of data attributes after each activity execution in the case.
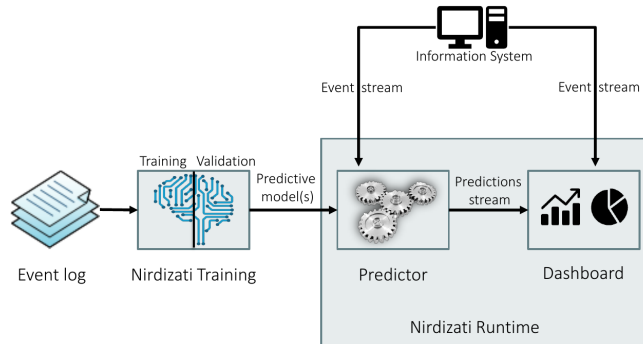
Fig. 1. High-level overview of Nirdizati.

As an example, consider a business process for managing loan applications; the applicant can be advised about the combinations of loan amount and duration that are the most likely to lead to an acceptance of the application, given the history of the application until now and the personal data of the applicant (e.g., age, salary, etc.).

Several approaches have been proposed in the literature to tackle common predictive process monitoring tasks. However, so far, these approaches have largely remained in the academic domain and have not been widely applied in real-time scenarios where users require a continuous predictive support.

In this paper, we present Nirdizati, a pioneering open-source web-based predictive monitoring tool, which is able to fill this gap between research and practice, by providing business analysts with a highly flexible instrument for the selection, generation and analysis of different predictive models, and end-users with continuous runtime predictions.

Nirdizati consists of two components: *Nirdizati Training* and *Nirdizati Runtime* (Fig. 1). Nirdizati Training takes as input a business process event log and produces one or more predictive models, which can then be deployed in Nirdizati Runtime. Once a model is deployed, Nirdizati Runtime listens to a stream of events coming from an information system supporting the process, and produces a stream of predictions. These predictions are then visualized in a continuously updated web dashboard.

2 Nirdizati Training

Nirdizati Training is the component of Nirdizati that allows users to produce predictive models later used by Nirdizati Runtime for making predictions on a stream of events. It provides several algorithms for generating predictive models suitable for different types of predictions and tailored to each specific dataset. For example, it is able to build predictive models for predicting remaining time, the next activity to be performed, whether a certain outcome will be achieved
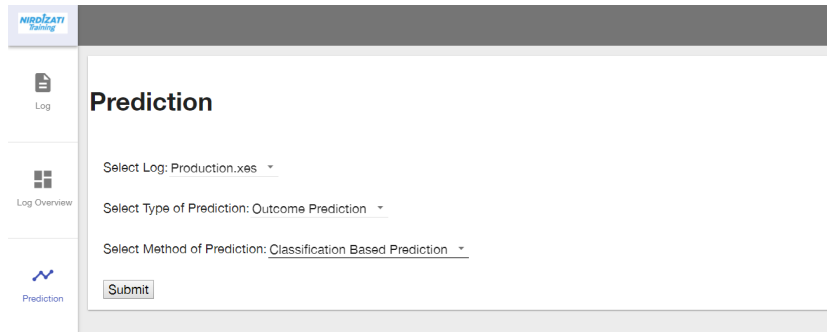
Fig. 2. Front-end of Nirdizati Training.

or not and the overall workload per day. To this aim, the training component of Nirdizati relies on two phases: a training and a validation phase. In the former, one or more predictive models are built; in the latter, their suitability to the specific dataset is evaluated, so as to support the user in selecting the predictive model that ensures the best results.

Nirdizati Training is composed of a front-end application (Fig. 2), which allows users to select the prediction methods and to assess the goodness-of-fit of the built models, and a back-end application responsible for the actual training and validation. The back-end application is, in turn, composed of four submodules shown in Fig. 3. A Log Manager is in charge of managing the logs. Uploading and retrieving the logs are the basic operations of this module. The Encoder is responsible for parsing the logs, labeling them according to the desired type of prediction, and preparing the data for the training phase. In the Training
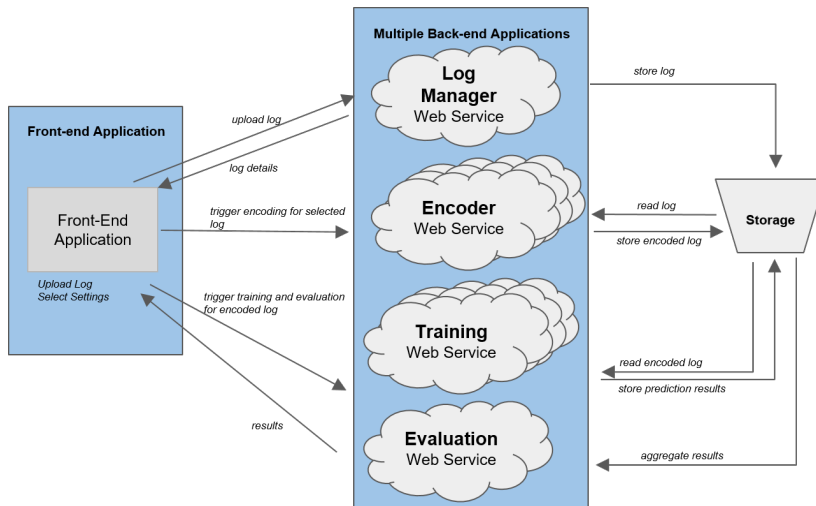

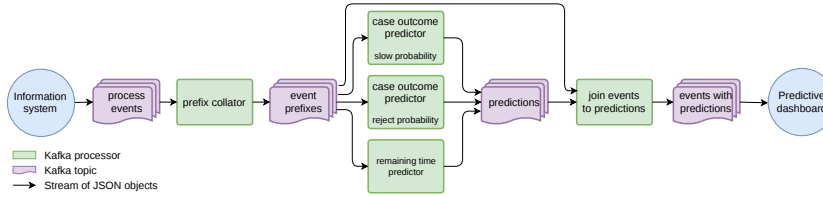**Fig. 3.** High-level overview of Nirdizati Training.

Fig. 4. High-level data flow diagram of Nirdizati Runtime.

submodule, the encoded data is split into training and validation set used for evaluation purposes, and the predictive models are built from the training data. Finally, the Evaluation submodule tests the validation data against the created model(s) to get accuracy measures with respect to the ground truth, available from the complete traces in the validation data. The back-end application comes with a storage module for saving the uploaded logs and the generated predictive models.

3   Nirdizati Runtime

Once the predictive models have been created, they are used by the Runtime component to make predictions on ongoing cases. Nirdizati Runtime takes a stream of events produced by an information system, transforms it into a stream of predictions, and visualizes those predictions in a web-based dashboard. The transformation is implemented using the dataflow pipeline in Fig. 4. The pipeline is built on top of the open-source Apache Kafka stream processing platform. The "predictor" components of the pipeline are the predictive models from Nirdizati Training. The "topic" components are network-accessible queues of JSON messages with publisher/subscriber support. This allows the computationally intense work of the predictors to be distributed across a cluster of networked computers, providing scalability and fault-tolerance. The "collator" component accumulates the sequence of events-to-date for each case, such that the prediction is a stateless function of the trained predictive model and of the case history. This statelessness is what allows the predictors to be freely duplicated and distributed. The "joiner" component composes the original events with the various predictions, ready for display on the dashboard.

The dashboard provides a list of both currently ongoing cases (colored in gray) as well as completed cases (colored in green), as shown in Fig. 5. For each case, it is also possible to visualize a range of summary statistics including the number of events in the case, its starting time and the time when the latest event in the case has occurred. For the ongoing cases, Nirdizati Runtime provides the predicted values of the performance indicators the user wants to predict. For completed cases, instead, it shows the actual values of the indicators. In addition to the table view, the dashboard offers other visualization options, such as pie charts for case outcomes and bar charts for case durations.
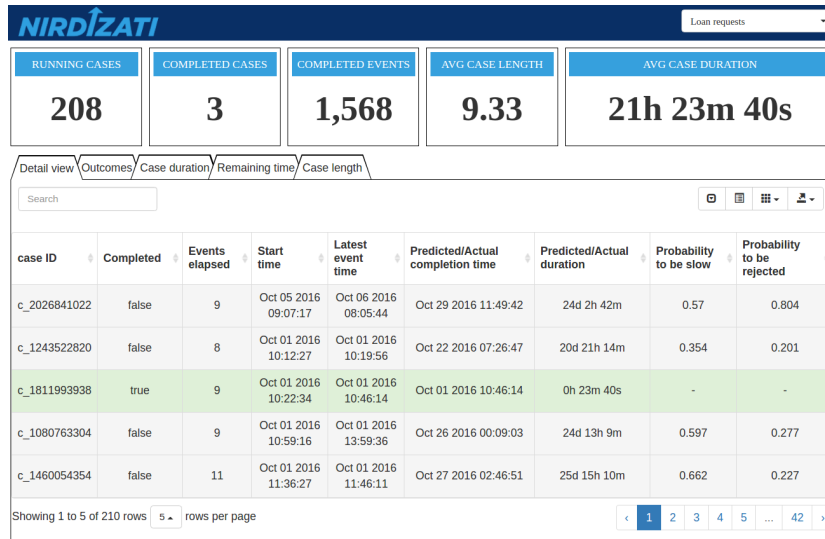
Fig. 5. Main view of Nirdizati Runtime.

Process workers and operational managers – typical users of Nirdizati – can set some process performance targets and subscribe to a stream of warnings and alerts generated whenever these targets are predicted to be violated. Thus, they will be capable of making informed, data-driven decisions to get a better control of the process executions. This is especially beneficial for processes where process participants have more leeway to make corrective actions (for example, in a lead management process).

4   Conclusion

Nirdizati is a configurable full-stack web application that supports users in selecting and tuning prediction methods from a list of implemented algorithms and that enables the continuous prediction of various performance indicators at runtime. The predictions are presented in a dashboard that offers multiple visualization options.

A video demo of Nirdizati can be found at `http://youtu.be/0nr14lX04-I`. A public release of Nirdizati Training and Nirdizati Runtime is available at `http://training.nirdizati.com` and `http://dashboard.nirdizati.com`, respectively. The source code is available under the Lesser GNU Public License (LGPL) at `http://github.com/nirdizati`.

References

1. Maggi, F.M., Di Francescomarino, C., Dumas, M., Ghidini, C.: Predictive monitoring of business processes. In: Advanced Information Systems Engineering - 26th International Conference, CAiSE 2014. Proceedings. pp. 457–472 (2014)