POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Executive Summary of the Thesis

Milan Court of Justice Immigration Section Processes Analysis

Laurea Magistrale in Management Engineering - Ingegneria Gestionale

Author: Gabriele Liuzzo

Advisor: Prof. Barbara Pernici

Academic year: 2022-2023

1. Introduction

Immigration is a very relevant social phenomenon: it involves millions of people all over the world and poses challenges and opportunities for both the countries of origin and those of destination. Immigration in Italy is a phenomenon of great social importance, the year 2023 registered one of the highest number of migrant arrivals on Italian territory ever, most of whom applied for international protection. Among the institutions responsible for managing and regulating migratory flows, the Courts play a fundamental role, called upon to rule on requests for asylum, citizenship, residence permits and expulsion of migrants. The most problematic part of Italian Justice system is its timeliness.

Italy suffers from some of the longest trial durations in Europe, especially in civil matters. In 2020 the average length of civil trials is 674 days for the first instance, more than twice the European average of 237 days. The second instance takes 1,026 days on average, almost six times longer than the European average of 177 days. The Court of Cassation requires 1,526 days on average, almost nine times longer than the European average of 172 days[1].

Reasonable durations of trials are important to respect the right to a fair trial and support the protection of the rights of citizens and businesses, which leads to greater trust in the justice system. The duration of trials depends on the nature, complexity and type of cases, and can be influenced by external factors, such as legislation, organisation and resources.

The reduction of juridical times is an objective on which the juridical system has been strongly focused to improve in recent years. Among the objectives of the National Recovery and Resilience Plan there is also a 40% reduction by June 2026 in the duration (measured on the three levels of judgment and according to the disposition time indicator) of civil trials and a 25% reduction in the duration of criminal proceedings [6].

The digital civil trial is a project of the Ministry of Justice, with the aim of improving the quality of the Italian legal system, speeding up trials and reducing its costs. Its functioning is based on a complex informatic architecture, where the SICID (acronym for District Civil Litigation Information System) is the database that records the history of all Court proceedings. Chancellors and magistrates have a direct access to it, being

[1] https://rm.coe.int/cepej-fiche-pays-2020-22-e-web/1680a86276

able to gradually register the events that compose each process to the appropriate registers. It is from the SICID that the data on which this thesis is based were obtained, through an SQL query executed by the competent office of the Court of Milan.
The data set used comprehends 264,633 events for 20,787 Immigration cases that started between January 2018 and October 2022. This thesis aims to offer a methodology that can be used to systematically analyze the specialized Immigration Section processes, offering an overview and techniques that could help to build a monitoring tool able to improve the management of the cases. This thesis also aims to verify the feasibility of a machine learning model suitable for predicting the duration of processes at their beginning.

2. Related Work

The extraction and analysis of the events data generated by IT systems have become fundamental elements for improving the organizations of companies and institutions over the last two decades. Process mining is a family of methods that combine process optimization with data analysis. It allows organizations to take full advantage of the information stored in their systems, it also used to verify process compliance, detect bottlenecks and predict execution problems [7]. Since this thesis is part of the larger project "Next Generation UPP: For the improvement of efficiency and of Justice Services in North-West Italy"[2], the work already carried out has served as a solid starting point. The thesis [2] was useful as an introduction as it analyzed and measured the timing of the Appeal Court proceedings and it also evaluates the median duration of the processes for each legal subject. [4] suggests a process mining approach that aims to break down the waiting time observed in each activity transition within a process into various direct causes. It also seeks to examine the influence of each identified cause on the overall efficiency of the process. In 2014, a study [5] was also developed which aims to detect how much the development of institutions in a certain territory attracts professionally qualified migration. The paper [3] in collaboration with the Court of Livorno, which provided the cases data, studied the application of the legal constraints and defined performance indicators capable of real time tracking (week by week) of the performance of chancellors and judges working in the Court. Focusing on the prediction of the duration of processes, several papers can be found with different approaches developed especially in recent years. In [9] using regression techniques, the time needed to close trials registered in Texas Courts over a considerable period of 17 years was estimated. In [8], trials in three different legal districts in Oregon were analyzed with the aim of comparing their performances according to the type of organization present within them, suggesting that greater autonomy and independence of the staff within the Court is correlated with greater speed in resolving disputes. This thesis finds space for work consisting in an in-depth analysis with process mining techniques of the specialized Immigration Section which is characterized by processes of a very particular nature compared to the processes of the rest of the Court from a legal point of view. Predictive machine learning models are also applied for the first time to Immigration Section data, evaluating the predictive importance of various factors and the feasibility of a predictive model that can anticipate the duration of the cases at their beginning.

3. Methodology

This section describes the methodology developed starting from the database extracted from the Court information system. The database is composed by several parts, each related to a different aspect of the Court processes. The development of a systematic methodology allows the extraction of useful information that can be used to monitor and improve the management and execution of processes. The possibility of creating a machine learning model capable of predicting the total duration of a new process at its beginning is also discussed, thus making it possible to anticipate cases that could prove critical. The software used to develop this methodology is Matlab, a software designed for matrix analyses distributed since 1984 by Mathworks[3]. It is a platform that integrates computation, visualization and programming.

[2] https://www.nextgenerationupp.unito.it

[3] https://it.mathworks.com/products/matlab.html

3.1. Processes Profiling

The starting data was transformed to permit their analysis, in this manuscript only the most critical steps are reported. The most important parts of the database are 'fasc' and 'stor'. The 'fasc' part contains a list of all cases and their characteristics, while the 'stor' part records the events history of each case. The first step was to filter the database extracted from the Court to select only the processes related to the Immigration Section. The filtering criteria was the column "ctipse", which specifies the section where the process is registered, equal to "si", the Immigration Section identifier code. The next step was to compute the process identification keys by concatenating the columns "cnuruo" and "canruo", which form a unique identifier for each process. The process ID allowed the integration of the different parts of the database into a single file. Many analyzes were carried out which can be grouped according to three different aspects: the first being the characteristics of the cases at their date of registration such as the object or 'theme' in which they are classified and their accumulation over time. Then the judges are analyzed, measuring for each one the probability to accept or deny a request of asylum and the assignments of cases for each one over time. And finally it is also statistically examined how the postponements of hearing affect the total of time that passes from the hearing scheduling date to its actual date.

3.2. Phases Analysis

This second analysis aims to divide the processes into four distinct phases (assignment of the judge, scheduling of the hearing, waiting for the hearing, and the decision phase) that represent the four most important parts of the processes. To this end, five key steps were identified and their corresponding informatics representations delimiting these stages were isolated. These steps are designed for immigration processes, their validity are not verified for other categories of processes, since they constitute a very specific type of processes. The dataset has been reduced selecting the trials that began in 2018 or 2019 resulting in 7,717 processes (the 78% of all concluded processes). This aims to mitigate a selection bias that would occur if all concluded process would be analyzed, since only the fastest ones have been concluded among all the process the recently started. The software Apromore[4] was used to generate graphs that visually illustrate the phases durations.

3.3. Cases Duration Prediction

Machine learning models constitute an important part of the broader field that can be defined as artificial intelligence. Their objective is to recognize and 'learning' patterns after having been trained on pre-existing data and subsequently being able of making predictions on new data. In this analysis a model to predict the entire duration of the processes at their beginnings is built and discussed. To develop the models, variables have been first selected and eventually computed. Specifically, the starting date of the trial, the judge assignment delay, the object or 'theme' of the process, the rituality (a legal set of rules that the process has to follow), the number of ongoing processes, the assigned judge, the clerk identifier that corresponds to the judge designation events, and how many sub-process the case is composed of. The three most important predictors are discussed in the results. 10% of the processes have been set selected in a heterogeneous manner to use them as test cases. To compare the different models the mean absolute error was used rather than the root mean square error which is generally taken as a reference. It has been preferred as it is less affected by extreme outliers, which in the context in question are due more to the individuality of the case than to the factors that the Immigration Section can control.

4. Results

4.1. Processes Profiling Results

Developing a methodology to profile processes characteristics, judges performance and hearing postponements is useful to gain an overview on how the entire Immigration Section works. The analysis carried out shows that cases beginning have been heavily concentrated in years 2018 and 2019 (13,178 cases, the 63% of all processes) and low amount of new arrivals in the following years (2020-2021-2022) visible in Fig. 1). Moreover, it has been discovered that the 17,049 cases (80% of the total) are evaluations of asylum re-

[4] https://apromore.com/

quests, while the other 20% are subdived in 17 different categories also relevant to immigration for a different reason. The probability of judges to accept or deny an asylum requests has been also analyzed: while all the three judges with more assigned cases share the same asylum requests acceptance rate around 40%, the others judges manifest great disparity up to 98% even though an interview with an expert would be necessary to be able to contextualize completely these statistics. Furthermore, it has been measured that hearings are scheduled with a median of 51 days to the scheduling date and an average of 66. The 31% of cases are postponed to an additional waiting of 105 days as median and 136 days as average.

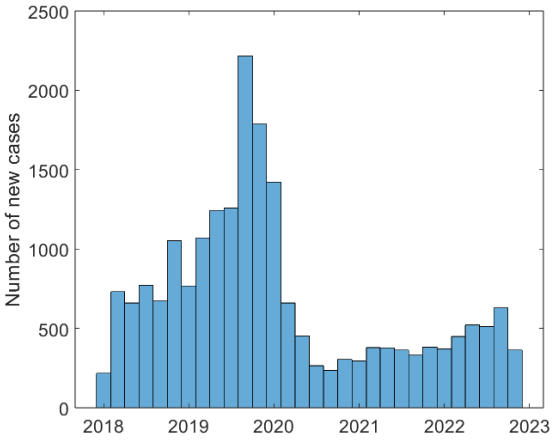

Figure 1: New cases for every two months

4.2. Phases Analysis Results

By isolating the crucial steps that compose the Immigration Section processes it has been possible to analyze the duration of the phases that these steps determine. The isolated steps are: beginning, designation of the judge, hearing scheduling, hearing, sentence. In the dataset available 20,762 processes started, 20,313 have been assigned to a judge, 12,015 had an hearing date scheduled, 10,975 had an hearing, 9,226 have been completed. Considering the 7,177 cases that started in 2018 or 2019 and have been also concluded, the duration of the phases that the steps delimit are as follow (median durations): 27 days to assign a judge (avg. of 38), 128 days to schedule an hearing (avg. of 242), 57 days to reach an hearing, (avg. of 107), 161 days to reach a sentence (avg. of 247). The total duration of the cases started in 2018 or 2019 that have been concluded by 2022 is 1.6 years by median and 1.72 years as average. The designation of the judge is the most fast and stable phase, the second and third phases have a good chance to be fast, but it is possible that a long period of time is need to complete them, the decision phase is the longest and unpredictable one.

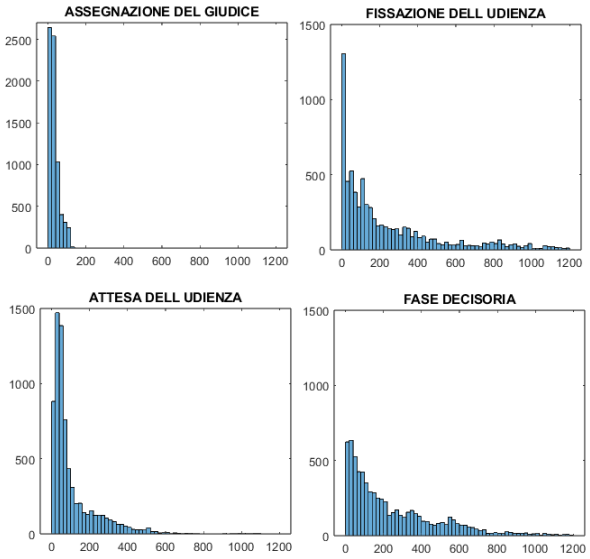

Figure 2: Phases histograms. X axis are days while Y axis are number of cases

4.3. Cases Duration Prediction Results

The three main variables that carry the most prediction value are: designed judge, object of the process, rituality of the process. Judges who have fewer cases tend to show a smaller possible range of variability, whereas judges who have many cases tend to show a more homogeneous distribution of durations. The judge who is responsible for the most cases shows that a significant part of them are completed within two years. Regarding the processes objects, the appeal to art. 35 the most frequent object is linked with wide variability: 50% of cases (second and third quartile) are between 408 and 941 days, the total range that the duration of these cases covers is larger than four years. However, the trials of other types of objects enjoy a narrower range: emergency measures cases last less than 100 days, while the cases with object 'revocation of the sentence' permanently lasts around 200 days. "Dublin Unit" appeals processes within the lower two quartiles reach sentences in less

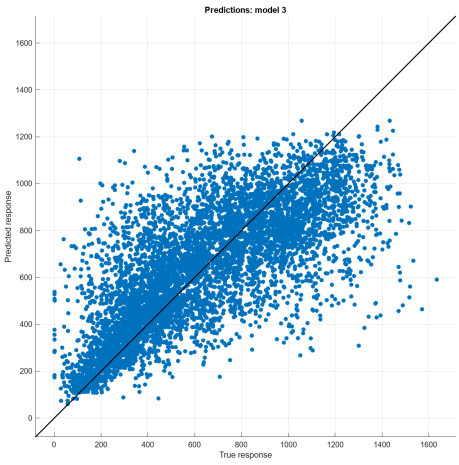

Figure 3: Predicted duration over true duration. Random forest model.

than 91 days. Ritualities, sets of rules that regulates the conduct of a trial, also has predictive power: the 7O has a rather limited duration compared to the others. The 1O, 5O, AC ritualities (set of juridical rules the the process has to comply to), predict a narrow duration range. The model produced with a decision tree has an average absolute error of 170 days. The cases that are particularly difficult to predict are the particularly long ones (more than 40 months), which the model tends to predict generally shorter in respect to the true duration. These are cases probably linked to the unpredictability of human behavior rather than a natural distribution of process durations. 167 days is a rather wide range, but considering the context that these durations are estimated at beginning of the processes and that they can last more than three years, it can be considered an interesting result. A Random Forest model was also trained, it is more computationally expensive but tends to offer more reliability. Like the decision tree, there is greater difficulty predicting the duration of longer cases. The mean absolute error decreases to 159 days. Fig. 3 shows the predicted duration against the true duration. The digitalisation of the urgency mark that is applied to the physical cases files would greatly improve the performance of the machine learning models.

5. Turin Immigration Section Comparison

Since this thesis is part of the larger project "Next Generation UPP: a project to improve the performance of justice in north-west Italy" it was possible to get in touch with a team from the Polytechnic of Turin that produced an unpublished manuscript entitled "Analysis of the Impact of UPP Employees in Improving the Performance of Judicial Offices: Study on the Labor and Immigration Sections of the Court of Turin". Two main points emerge from the comparison: the first being that the digitalisation of the country of origin but also the digital equivalent of the "urgent" mark which is applied to the physical files of the cases would be extremely useful, to let workers be more efficient and to achieve better analysis and prediction models. The second point regards the AUPP staff: with the recruitment announced by the Legislative Decree of 9 June 2021 [1], around 8,000 law graduates were introduced in the Courts to increase the speed and efficiency of the Sections. The work of the Polytechnic of Turin detected the impact of the new personnel to be overall positive and useful with a high degree of tasks specialization. In a meeting with the Head of the Chancellery Office of the Milan Immigration Section instead doubts were expressed about the desired beneficial effect: some of the work that the new staff produced was often not compliant with the standards and therefore required additional work and that the Milan staff have been employed without specialized tasks assigned. As the team from the Polytechnic of Turin reports, "In the Immigration Sections, throughout the national territory, there is a greater problem linked to the training of UPP Employees than the training of the same figures who operate in the other Sections of a Judicial Office: in the latter, in fact, there is no need to explain the different legal institutions to law graduates, but only to show their tasks. In the Immigration Sections, however, the UPP staff must also be trained on the subject itself". From the comparison of the two different effects that the UPP employees have on the two Immigration Sections, it can be deduced that the usefulness of the inclusion of new personnel in the juridical system is strongly improved by the training that is dedicated to them first, that helps to perform high

specialization tasks.

6. Concluding remarks

In this work, a methodology has been proposed to analyze the data that the Immigration Section of the Court of Milan generates and stores on the SICID information system. The developed methodology consists of replicable analyzes which, starting from the raw data extracted directly from the information system, achieves the measurement of processes phases duration and other statistics useful for monitoring and therefore improving the efficiency of the Immigration Section. The feasibility of machine learning models applied to predicting cases duration is also discussed and points emerged from the parallel work conducted by the Polytechnic of Turin have been highlighted. A natural extension of this work would be to enrich it with additional cases, extracting again the data that the Court's information system has available. It would be critical that such data not only contain more useful parameters, but cover a wider time span. Since immigration cases can last up to four years, a dataset that covers at least eight years would be very useful to improve the reliability of the statistics collected, and to measure how the management of the cases changes over the years. From March 2023, the Cartabia reform abolished the rituality "7O - Summary Cognition 702 bis cpc" establishing the new ritual "art. 281 decies" as substitute. This new ritual consists as a simplified procedure of cognition which promises to be faster and efficient than the 7O rituality. A new data extraction from the Court's information system would allow access to the processes that began to benefit from it since March 2023. In addition, the digitalisation of the mark that is applied to the files that notifies their urgency, would make it possible to measure their impact on the processes carried out in parallel: how their hearings are rescheduled, how the decision phase is protracted, but also how the quality of the outcomes may vary. Moreover, there are 17 Immigration Sections operating in Italy, the comparison through the data that is automatically recorded in the systems could bring to light some interesting comparisons that can help mutual improvement.

References

[1] Consiglio dei Ministri. Decreto-Legge convertito con modificazioni dalla L. 6 agosto 2021. *Gazzetta Ufficiale della Repubblica Italiana*, 2021.

[2] Mattia Del Treste. Temporal Analysis of Italian Civil Cases: the Appeal Court of Milan. *Master Thesis in Managment Enginnering - Polytechnic of Milan*, 2022.

[3] Beniamino Di Martino, Luigi Colucci Cante, Antonio Esposito, Pietro Lupi, and Massimo Orlando. Supporting the Optimization of Temporal Key Performance Indicators of Italian Courts of Justice with OLAP Techniques. *Complex, Intelligent and Software Intensive Systems: Proceedings of the 15th International Conference on Complex, Intelligent and Software Intensive Systems (CISIS-2021)*, 2021.

[4] Katsiaryna Lashkevich, Fredrik Milani, David Chapela-Campa, Ihar Suvorau, and Marlon Dumas. Why am I Waiting? Data-Driven Analysis of Waiting Times in Business Processes. *International Conference on Advanced Information Systems Engineering*, 2023.

[5] Annamaria Nifo and Gaetano Vecchione. Do Institutions Play a Role in Skilled Migration? The Case of Italy. *Regional Studies Volume 48*, 2014.

[6] Uffici del Senato della Repubblica e della Camera dei deputati. Piano Nazionale di Ripresa e Resilienza. , 2021.

[7] Wil van der Aalst. Process Modeling and Analysis. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, 2011.

[8] Wendy A Walsh, Tonya Lippert, Meredyth Goldberg Edelson, and Lisa M Jones. Length of Time to Resolve Criminal Charges of Child Sexual Abuse: A Three-County Case Study. *Behavioral Sciences and the Law*, 33, 2015.

[9] Jun Drs Zhou. Determinants Of Delay In Litigation: Evidence and Theory. *American Law and Economics Association Annual Meetings*, 2008.