Survival analysis as semi-supervised multi-target regression for time-to-employment prediction using oblique predictive clustering trees

Viktor Andonovikj [a,b,*], Pavle Boškoski [a], Sašo Džeroski [a], Biljana Mileva Boshkoska [a,c]

[a] *Jožef Stefan Institute, Jamova cesta 39, 1000 Ljubljana, Slovenia*
[b] *Jožef Stefan International Postgraduate School, Jamova cesta 39, 1000 Ljubljana, Slovenia*
[c] *Faculty of Information Studies in Novo mesto, Ljubljanska cesta 31b, 8000 Novo mesto, Slovenia*

ARTICLE INFO

ABSTRACT

We address the problem of estimating the time-to-employment of a jobseeker using survival analysis and oblique predictive clustering tree. Unlike standard survival analysis, oblique predictive clustering tree can handle categorical and continuous data and is capable of modelling non-linear dependences. Treating the censored data as missing data opens the possibility to perform survival analysis by using structured output prediction in semi-supervised multi-target regression setting. The effectiveness of this approach is shown on a real dataset from Public Employment Services in Slovenia, comprising time-to-employment records with jobseekers' personal and professional characteristics. The performances are compared with six state-of-the-art AI methods. To the best of our knowledge, this is the first example of using semi-supervised oblique predictive clustering tree for survival analysis.

1. Introduction

Developing data-driven systems addressing unemployment issues has significant practical merit for all Public Employment Services (PES). On daily basis, PES, deals with an increased in-flow of high diversity and fast changing skills of jobseekers, while counteracting the risks of widespread long-term unemployment. Therefore, a reliable time-to-employment estimate of a jobseeker is a precious information (Yizhong et al., 2017). This paper aims to provide a probabilistic and computationally efficient time-to-employment estimation model.

A time-to-event estimate is usually obtained by using techniques from survival analysis (Anon, 2016), especially when dealing with partially observed or truncated data. Time-limited observations represent one class of such data, including human response to a particular medical or a health treatment, a death in biological organisms (Yang et al., 2011), failure in mechanical systems (Pereira et al., 2018), and among others — time-to-employment. Survival models can provide a proxy assessment of the effectiveness of particular actions, for instance, medical treatments. Although such models can be used for determining important features, in many cases, they impede a positive outcome or may lead to segregation and profiling of people, which may be considered as undesired characteristics of a model describing a social situation. Hence, the usage of standard survival models is socially questionable, since the outcome of these analyses typically has significant practical consequences.

Another issue is that standard methods in survival analysis, such as the Cox proportional hazards model (Cox, 1972), impose restrictive assumptions. One central assumption is that the relative hazard rate remains constant over time with different covariate levels or predictors (Babińska et al., 2015). On the other hand, Machine Learning (ML) methods are only partially suitable for time-to-event analysis, primarily due to censored data. We bridge the issue of handling censored data by proposing to learn a survival model with semi-supervised learning. With learning of semi-supervised predictive models, for instance predictive clustering trees (PCTs), at our disposal the problem of time-to-employment estimation can be transformed into a problem of semi-supervised multi-target regression (Roy, Stepišnik, Vens, & Džeroski, 2022). The main disadvantage of semi-supervised learning of PCTs is their computational complexity. They scale poorly with high-dimensional input. Hence, to address the scalability problem of learning PCTs, we propose a solution based on semi-supervised oblique predictive clustering trees (SPYCTs). Compared to the axis-parallel trees, which search for the best split of a dataset with respect to a single

---

feature, the oblique trees use all the features to find the optimal node split. The use of oblique trees for survival problems has already been proposed (Kretowska, 2020). The authors in that paper propose three different recursive partitioning methods to cope with discrete survival data. In their approach, the time points are grouped into time-intervals for which the models make predictions of survival probability. The issue is that for each of those intervals uncensored examples are required, the more the better — which restricts the number of time-intervals to be low, in order for the model to perform well. Also, the method yielding the best predictive performance is based on local, independent trees used for single-target prediction at each time interval.

Global approaches, where one model is used to simultaneously predict the output for all time points, are proven to perform at least equally well as the local approaches. Global approaches are more efficient and need a smaller number of trees in an ensemble to achieve optimal performance (Kocev, Vens, Struyf, & Džeroski, 2013). Additionally, in the local models, the parameters for prediction at different time points are not shared between the models, meaning that the problems are treated as independent, which in our case, would be a false assumption.

The aim is to show the effectiveness of the global approach which uses semi-supervised SPYCT and to show that it is possible to interpret the predictions of such a model — meaning to be able to explain based on which features the model comes to the final decision.

The use of semi-supervised models to estimate the values of a survival function was already proposed by (Roy et al., 2022). However, in that paper, the authors proposed the use of semi-supervised PCTs, which scale poorly with higher input and (especially) output dimension. This makes it challenging to use semi-supervised PCTs in large-scale settings. We address this issue by proposing the use of the more efficient semi-supervised SPYCTs in the same framework. This makes it possible to use this framework in large-scale settings. We also show that it is possible to interpret the predictions of the model with respect to the input features, which is another valuable contribution. Our approach contributes to the field of survival analysis by providing a powerful and efficient approach that handles categorical data, models non-linear dependencies, addresses censoring, and offers interpretability. The combination of these advantages distinguishes our method from existing approaches and highlights its potential impact on various domains requiring accurate and interpretable survival predictions.

The remainder of the paper is organised as follows. Section 2 provides an overview of the current implementations and research of unemployment modelling for labour market analysis. Section 3 provides a brief background on survival analysis and SPYCTs. Section 4 describes the methodology of transforming the task of survival analysis into a semi-supervised multi-target regression framework, and the procedure of making a prediction using a semi-supervised SPYCTs model. Section 5 gives an overview of the dataset we use and its features that are used for modelling, the experimental setup, and the evaluation metric. Section 6 presents the performance evaluation results of the models, and a description of the methodology for explaining the predictions of semi-supervised SPYCT. Finally, we summarise the paper's key takeaways in Section 7.

2. Related work

2.1. ML-based models for survival data

The main challenge in using ML approaches for survival analysis is that it is challenging to deal with censored data. Consequently, several extensions of classical ML approaches have been developed for survival analysis cases dealing with censored data (Wang, Li, & Reddy, 2017). ML-based survival analysis approaches can be roughly categorised into two groups. The first one represents methods that model the hazard function as a regression problem and then use it to estimate the survival curve. The second group consists of models that directly estimate the survival curve given the input data.

Models based on the hazard function. A typical example of this category is the so-called DeepSurv model (Katzman et al., 2018), a Cox proportional hazards deep neural network. The DeepHit model (Lee, Zame, Yoon, & der Schaar, 2018) directly estimates the joint distribution of events in discrete time. A neural variant of the classical Multi-Task Logistic Regression (MTLR) has also been proposed (Fotso, 2018). While, in the classical MTLR variant, the relations between the features are modelled with a linear function, a multi-layer perceptron is used in the neural network variant, which can capture non-linearities. The PCHazard model (Kvamme & Borgan, 2021) assumes that the continuous-time hazard function is constant in some predefined intervals. For smaller datasets, this approach leads to improved predictive performance.

Models which directly estimate the survival curve. The second group of methods directly estimate the survival curve from the input data. These methods follow the paradigm of structured output prediction (SOP), and the output is usually a vector of numerical values (probabilities) made by either local or global models. Local approaches use a separate model for prediction at each discrete time-point of the survival curve (Kretowska, 2020). Global approaches use a single model to predict the whole vector of probabilities of the event happening at each discrete time-point (Roy et al., 2022).

New methods related to clustering and community detection have been recently proposed. One approach proposes characterising the fuzzy community structure via likelihood optimisation (Li et al., 2022). Another approach proposes the use of a fast Markov clustering algorithm based on belief dynamics (Li, Xu, Qiu, & Pei, 2023). Such methods for detecting fuzzy communities or cluster configurations in complex networks can provide insights into complex community structures and inform predictive models to enhance their accuracy. Methods based on utility mining for sequence data have also been proposed (Zhang, Du, Yang, Gan, & Yu, 2021). Such methods are advantageous for predictive modelling by effectively extracting on-shelf high-utility sequential patterns, reducing search space, and improving efficiency, thereby enhancing the accuracy and performance of predictions.

None of the above methods resolves the requirements of explainability completely. We address this issue by proposing the use of semi-supervised SPYCT for survival data and a methodology to interpret the predictions made by the model. The model belongs to the group of ML-based methods that directly estimate the survival curve from the input data.

2.2. Probability of exit (PEX) models

The first practical implementations of predictive models for time-to-employment estimation appear in the period of 1990–2000 (Kureková et al., 2014; Niklas, Sztandar-Sztanderska, Szymielewicz, Baczko-Dombi, & Walkowiak, 2015). There are roughly four distinct groups of methods through which time-to-employment of a jobseeker can be estimated. The first group consists of logit or probit-based approaches (O'Connell, McGuinness, Kelly, & Walsh, 2009; Riipinen, 2011; Wijnhoven & Havinga, 2014). The second group represent methods whose goal is to estimate the probabilities of going in and out of the labour market (Riipinen, 2011; Sengul, 2017; Shimer, 2012). The third group consists of methods based on ML (Andonovikj, Boškoski, & Boshkoska, 2021; Ballot, Kant, & Goudet, 2016; Boškoski, Perne, Rameša, & Boshkoska, 2021). The final group consists of methods based on labour flow networks (Andonovikj, Boskoski, Evkoski, Redek, & Boshkoska, 2022).

The drawbacks of the current implementations include weak predictive performance, no uncertainty estimation, computational inefficiency. They also include producing non-explainable or black-box models. All of the above-mentioned implementations have at least one of these drawbacks.

Our method belongs to the group of methods based on ML. It addresses the main challenges in the current implementations — the

ability to handle vast amount of heterogeneous data, uncertainty estimation (in the form of survival curves), and the explainability behind the predictions of the model. To our best knowledge, it is among the first data-driven, but explainable ML approaches applied to model probability of exit (PEX).

3. Preliminaries

3.1. Background on survival analysis

Survival analysis deals with time-to-event data that is not always fully observed. This means some cause for censoring exists, which interrupts the observation before the actual event occurs. Right-censoring is the most common type of censoring, and it occurs when the knowledge of survival is incomplete on the right side (towards the end) of the follow-up period. Consequently, each observed survival record consists of three variables: the actual survival time T, the observed portion of T denoted by t, and a discrete indicator δ, defined as follows:

$$\delta = \begin{cases} 0, & T > t \text{ right censored record.} \\ 1, & T = t, \text{ i.e., the event is observed and the record is not right censored.} \end{cases} \tag{1}$$

Furthermore, let us assume that each data sample can be represented with a set of covariates (features) $\mathbf{X}_i = \{x_{i_1}, \dots, x_{i_N}\}$. Therefore each data record i consists of $\mathbf{X}_i$, t_i and δ_i as in (1). It should be noted that, besides the more common right-censoring, there are two additional cases of censoring, i.e., left-censoring and interval-censoring, which are not of interest in the current analysis.

Let us assume that the distribution function of event times is $f(t)$. Based on these concepts, the focus of our analysis is on calculating the probability that the time of event occurrence T will be after a specific time t, i.e.

$$Pr(T > t) = S(t) = 1 - \int_0^t f(s)ds, \tag{2}$$

hence the name survival function $S(t)$. The survival function can be expressed also through the concepts of hazard rate $\lambda(t)$, which is defined as the probability that an event will happen in an imminent future condition, given that such an event did not already happen:

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} Pr(t + \Delta t | T \geq t) = \frac{f(t)}{S(t)} \tag{3}$$

The majority of the developed methods investigate continuous-time models, but there are also developments on discrete-time approaches (Roy et al., 2022). In essence, the methods described in Section 2 focus on (2) either directly or indirectly by modelling (3). The complexity of these two functions is where the ML approaches excel providing sufficiently high amount of quality data is available.

3.2. Oblique predictive clustering trees

PCTs are a generalisation of standard decision trees used for various SOP problems. PCTs are axis-parallel trees, meaning they only use single features in the splits and form axis-parallel hyperplanes. One of the issues with PCTs is that they scale poorly with output space dimensionality (Stepišnik & Kocev, 2020). In semi-supervised PCTs, all inputs are also considered as outputs. So they scale poorly also with input space dimensionality. As an alternative, SPYCTs were developed. SPYCTs keep the flexibility of PCTs, but use oblique splits. Unlike axis-parallel splits, oblique splits use a linear combination of all the features to split the dataset into two partitions. The splits correspond to a hyperplane in the input space. Oblique splits are appealing when decision boundaries are not aligned with feature axes. Besides axis-parallel and oblique splits, there is also the possibility of using non-linear splits (Ittner & Michael, 1996) that use a non-linear combination of the features to split the dataset. However, the non-linear splits are sensitive to noise in the data (Li, Dong, & Kothari, 2005) and are rarely used in practice. In contrast to PCTs, the interpretation of SPYCTs is not straightforward. Thus in the following sections, we describe a way of visually interpreting a SPYCT.

Let us assume that we have set of observations, and each is described with a feature vector $\mathbf{x} \in \mathbb{R}^N$. The observations can be divided into a subset $\mathbf{X}^{(l)} \in \mathbb{R}^{L \times N}$ of fully labelled observations and a subset $\mathbf{X}^{(u)} \in \mathbb{R}^{U \times N}$ of partially unlabelled observations. Partially unlabelled observations correspond to censored records and contain missing values after the time point of censoring. The complete feature matrix is $\mathbf{X} = \left[(\mathbf{X}^{(l)})^T (\mathbf{X}^{(u)})^T \right]^T \in \mathbb{R}^{(L+U) \times N}$. Similarly, the matrix containing the complete T targets for the L labelled observations is $\mathbf{Y}^{(l)} \in \mathbb{R}^{L \times T}$, and $\mathbf{Y}^{(u)} \in \mathbb{R}^{U \times T}$ where $\mathbf{Y}^{(u)}$ contains incomplete targets. In the context of multi-target regression, the word target/targets refers to the variable/variables that we aim to predict or estimate based on the input features. Therefore, the complete dataset D can be described as follows:

$$D = \left\{ \begin{bmatrix} \mathbf{X}^{(1)} \\ \vdots \\ \mathbf{X}^{(L)} \\ \mathbf{X}^{(L+1)} \\ \vdots \\ \mathbf{X}^{(L+U)} \end{bmatrix}, \begin{bmatrix} \mathbf{y}^{(1)} \\ \vdots \\ \mathbf{y}^{(L)} \\ \mathbf{y}^{(L+1)} \\ \vdots \\ \mathbf{y}^{(L+U)} \end{bmatrix} \right\} \tag{4}$$

The aim is to build a tree structure like in Fig. 1 that captures the information in (4) as close as possible. Each split node i of a SPYCT contains a fuzzy membership indicator (Stepišnik & Kocev, 2021), defined as:

$$FM_i = \sigma(\mathbf{x}^T \mathbf{w}_i + b_i) \in [0, 1], \tag{5}$$

where $\mathbf{w}$ and b are weights and bias, and σ is the sigmoid function defined as:

$$\sigma(\mathbf{x}) = \frac{1}{1 + e^{-\mathbf{x}^T \mathbf{w}_i + b_i}} \tag{6}$$

. For a particular split node i, the value FM_i tells us how much the corresponding instance belongs to the "left" group, whereas the value $1 - FM_i$ tells us how much it belongs to the "right" group.

The goal is to tune the parameters w_i and b_i $\forall i$ w.r.t. a split fitness function. Typically the split fitness function has the following form:

$$f(\mathbf{w_i}, b_i) = Z \cdot imp(FM) + (L + U - Z) \cdot imp(1 - FM), \tag{7}$$

where $imp(FM)$ is the impurity of the "right" group (Stepišnik & Kocev, 2020), calculated as:

$$imp(FM) = \sum_{k=1}^{N} \sigma_{\bar{X}_k}^2 + \sum_{k=1}^{T} \sigma_{\bar{Y}_k}^2. \tag{8}$$

The terms Z and $L + U - Z$ represent the sizes of the right and left subsets and are added to guide the split search towards balanced splits. The parameters $\sigma_{\bar{X}_k}$ and $\sigma_{\bar{Y}_k}$ are variances defined as:

$$\sigma_{\bar{\mathbf{X}}_k}^2 = \frac{1}{L+U} \sum_{i=1}^{L+U} \left(\mathbf{X}_k^{(i)} - \bar{\mathbf{X}}_{k,\mathbf{w}} \right)^2, \tag{9}$$

$$\sigma_{\bar{\mathbf{Y}}_k}^2 = \frac{1}{L+U} \sum_{i=1}^{L+U} \left(\mathbf{Y}_k^{(i)} - \bar{\mathbf{Y}}_{k,\mathbf{w}} \right)^2, \tag{10}$$

where $\bar{\mathbf{X}}_k$, and $\bar{\mathbf{Y}}_k$ are weighted averages defined as:

$$\bar{\mathbf{X}}_{k,\mathbf{w}} = \frac{w_k}{L+U} \sum_{i=1}^{L+U} x_k^{(i)}, \tag{11}$$

$$\bar{\mathbf{Y}}_{k,\mathbf{w}} = \frac{w_k}{J} \sum_{i=1}^{N} \bar{y}_k^{(i)}, \tag{12}$$

where J is the number of instances with known value at the kth column, and $\bar{y}_k^{(i)}$ denotes a known value of y_k. To get the impurity of the
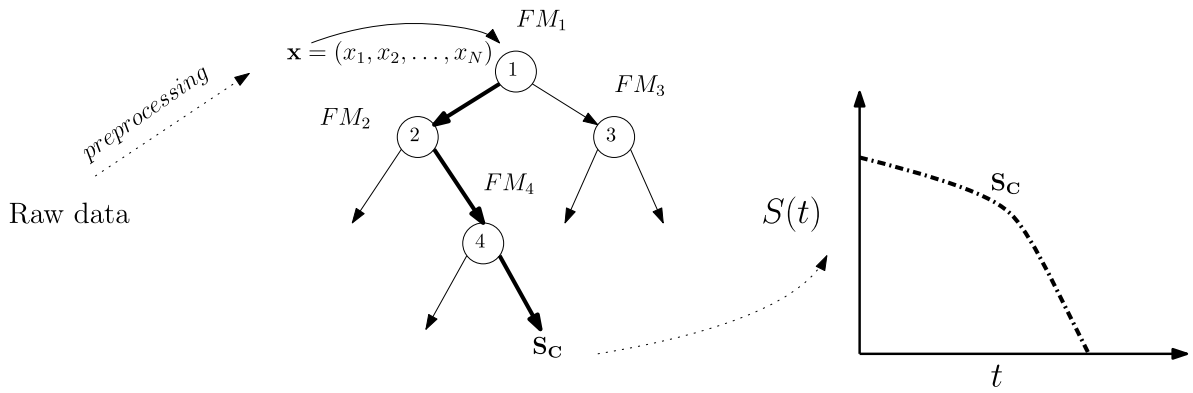
Fig. 1. Flowchart of the proposed framework. The process of making a prediction for a particular jobseeker starts with preprocessing the raw data. The preprocessed vector x is used as an input to the SPYCT model. Each of the non-leaf nodes contains an optimal fuzzy membership indicator FM learned during the training phase, which guides the split. The leaves contain a prototype used for making a prediction. The bold lines in the tree highlight the particular path for the example to reach a leaf node. The survival curve plotted corresponds to the prototype stored in the leaf node S_c.

"left" group $imp(1 - FM, p)$, the fuzzy membership weights are swapped with $1 - FM$ in (7).

The goal is to tune the parameters of each fuzzy membership function in order to optimally split the dataset into two partitions i.e. to minimise the split fitness function (7). The result is the following optimisation problem (Stepišnik & Kocev, 2021):

$$\min_{\mathbf{w_i}, b_i} \left(\|\mathbf{w_i}\|_{\frac{1}{2}} + C f(\mathbf{w_i}, b_i) \right) \tag{13}$$

where C is a constant intended to control the strength of regularisation. The function in (13) is differentiable, and is optimised using the ADAM optimiser (Kingma & Ba, 2014). The parameters of the ADAM optimiser β_1 and β_2 are set to their default values in PyTorch (0.9 and 0.99, respectively), and the learning rate ϵ is taken as a hyperparameter whose value is further determined through a process of hyperparameter selection, described in one of the following sections.

The process of SPYCT construction terminates upon the attainment of at least one of the following stopping criteria. One criterion is the minimum number of required examples in terminal nodes, with the requirement that there must always be at least one fully labelled example present to allow for the prediction of outcomes. Another criterion is the maximum depth d of the tree. If there is no test that significantly reduces the target variable variance — a leaf is created. The value of the vector in a single leaf (further used for prediction) is the mean of all the target variable values of the instances stored in that leaf.

When constructing variants of semi-supervised PCTs and SPYCTs, the clustering attributes K, used in the variance, consist of both features and targets. Therefore, $K = N + T$, where N is the number of features and T is the number of targets. The time complexity for learning a split in standard PCTs is thus $\mathcal{O}(DN \log N + NDK)$. The corresponding time complexity in semi-supervised SPYCTs is $\mathcal{O}(NI_o(D + K))$, where I_o is the number of optimisation iterations (Stepišnik & Kocev, 2021). PCTs scale quadratically with the number of attributes, while the gradient variant of the SPYCTs scales linearly (Stepišnik & Kocev, 2021). SPYCTs also gain an advantage by performing operations with sparse matrices. Thus, SPYCTs are much more computationally efficient, and this is one of the main motivations for using them in our use case, instead of PCTs.

4. Survival analysis as semi-supervised multi-target regression task

The survival analysis problem addressed within this paper is formulated as a semi-supervised multi-target regression problem. The survival curves addressing the time-to-employment of a jobseeker can be represented with a survival vector of target variables. Survival vector refers to a vector that contains survival probabilities for each time

point within a specific timeframe, in our case — for each day in the following 365 days. Thus the problem can be transformed to a semi-supervised multi-target regression task. In multi-target regression, the task is to predict multiple numeric variables. Our model outputs a vector of probabilities for a particular jobseeker to remain unemployed until each day of the 365 days following the initial enrolment in PES. For training purposes, we construct the target vector so that we include a binary value at each timestamp indicating that the jobseeker is still unemployed or not at that timestamp. The cases where the jobseeker has left the study, meaning PES has lost their track, are considered censored examples, which at the time points successive to that event have missing values in the target vector. If $\delta_k = 1$, the event occurred at time t_k and $y_{k_j} = 0$ for $j \geq k$. If $\delta_k = 0$, we have lost track of that instance, and we do not know what happened to it from timestamp t_k onward, therefore y_{k_j} are missing values for $j \geq t_k$. In such a way, the censored examples give rise to missing data, suitable for semi-supervised approaches. Our final model is an ensemble of 50 SPYCTs. Each of the leaves of the SPYCT holds a prototype (survival vector). Finally, we compare the predictive performance of the proposed approach to state-of-the-art survival analysis methods in terms of Brier Score (BS). Fig. 2 shows the schema of making predictions in the leaves of a SPYCT.

Making a prediction. For each new instance represented with feature vector $\mathbf{X}$, the procedure for making a new prediction starts by calculating the value of the fuzzy membership indicator (5) in the root node of the SPYCT. If the FM value is greater than $1 - FM$, we repeat the process in the "right" subtree. This procedure is recursively repeated until the model reaches a leaf node, where it makes the prediction. In our problem setting, the predictions from the model represent estimations of the survival function values:

$$\hat{y} \approx P(y_{ij} = 1 | \mathbf{X}_i) = P(T \geq t_j | \mathbf{X}_i) = S_i(t_j) \tag{14}$$

5. Dataset and experimental setup

5.1. Dataset

The dataset we consider consists of 74,086 instances and contains anonymised personal and professional characteristics of jobseekers that have entered the Slovenian PES. The dataset is complex because its attributes come in different forms (categorical, numerical, date and time), and most have to undergo transformation. An overview of the dataset and its features is given in Table 1.

The general structure of the jobseekers' attributes is described by dividing the attributes into several prominent groups: socioeconomic variables (gender, age), information on job readiness (education, health
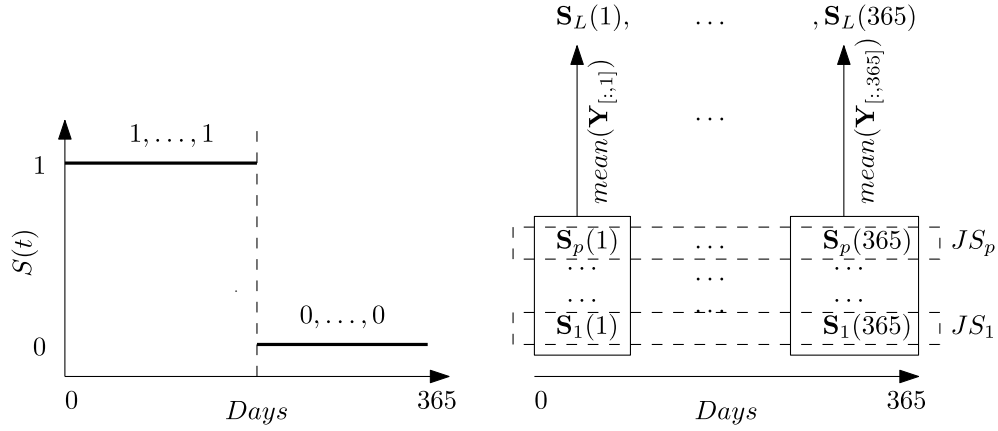
Fig. 2. Schema of the events in the leaves of a SPYCT. On the left hand side is an example of how the survival time of a jobseeker is represented as a vector before the training of the model. The vertical dashed line is the point where the jobseeker got employed. On the right hand side are the jobseekers (one per row) clustered in a particular leaf where $S_p(t)$ is the survival vector of example jobseeker p. The survival vector of the leaf $S_L(t)$, further used as a prediction, is calculated as the column-wise mean of the survival vectors of all the jobseekers clustered in that leaf.
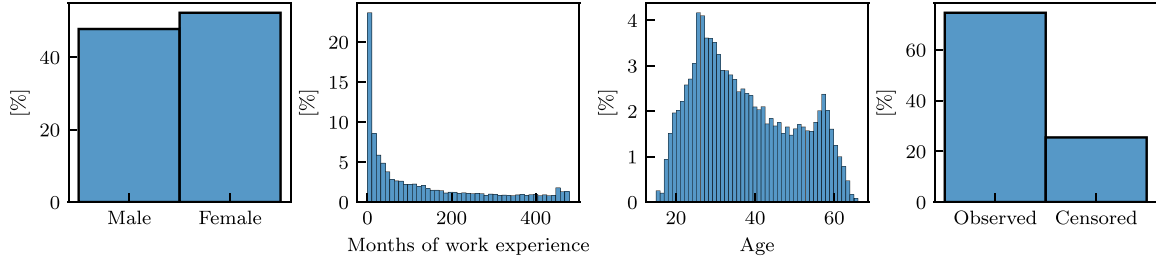
Fig. 3. Distributions of some of the dataset features.

Table 1
Features of the dataset and their cardinalities.

Continuous features	Discrete features
Day of PES entry	Education direction (30)
Months of work experience	Occupation (ISCO) (123)
Age	Reason for PES entry (5)
Month of PES entry	Reason for termination (43)
	Social benefits (2)
	Gender (2)
	eApplication (2)
	Unemployment benefits (2)

limitations), and opportunities (regional labour market development), and all available labour market history information, such as prior work experience. The target variable is in numeric form. It is a counter of days till a jobseeker gets employed or gets out of the study and leaves no further information about the jobseeker's employment status.

Fig. 3 shows the distributions of some features in the dataset. The sample of jobseekers is balanced in terms of gender. Most jobseekers enter the PES without any working experience. As expected, the age distribution is between 18 and 64 since this is the legal age interval of the labour-active population. It should be noted that the age distribution is bi-modal. The first peak is around 23–26 years since this is the expected graduation age for a university degree. The second peak is just before 60 years of age, i.e. jobseekers who became unemployed just before retirement due to various reasons. The final histogram shows that the dataset contains 30% of right censored data.

The categorical features are diverse with respect to their cardinality. The categories of the features with high cardinality (>1000 different class values per feature) are transformed using domain knowledge and the number of class values is significantly reduced. Because the *ISCO*

standard and *Education direction* provide a hierarchical classification of occupations and education directions, we took the higher-level classifications and thus significantly reduced the number of categories. The remaining categorical features were encoded using one-hot encoding. The temporal (cyclical) features, such as "Day of PES entry" and "Month of PES entry" were transformed by calculating the feature's sin and cos components. In this way, the artificially significant difference between, for example, the last day of the previous month and the first day of the current month is removed, and the value of the feature can be viewed as a point $(\sin(x), \cos(x))$ of a circle.

5.2. SPYCT and its hyperparameter values

Our model is an ensemble of SPYCTs. The best set of hyperparameter values for the particular use case was obtained through grid search method. The hyperparameters that were taken into account in the grid search were:

- the number of trees in the ensemble $NT \in \{1, 50, 100, 150, 200, 250, 500\}$
- the maximum depth of a tree $d \in \{3, 4, 5, 7, 10\}$, and
- the learning rate of the ADAM optimiser $\epsilon \in \{0.001, 0.005, 0.01, 0.05, 0.1\}$.
- regularisation constant $C \in \{2.5, 5.0, 7.5, 10.0, 12.5, 15.0\}$

The objective function utilised to select the best values of the hyperparameters for the particular use case was the average Brier score over all the discrete time points. The following values were obtained for the SPYCT hyperparameters: $NT = 50$, $d = 3$, $\epsilon = 0.1$, and $C = 10.0$.

5.3. Competing methods

The performance of the proposed approach is compared with the performance of the following survival analysis methods:
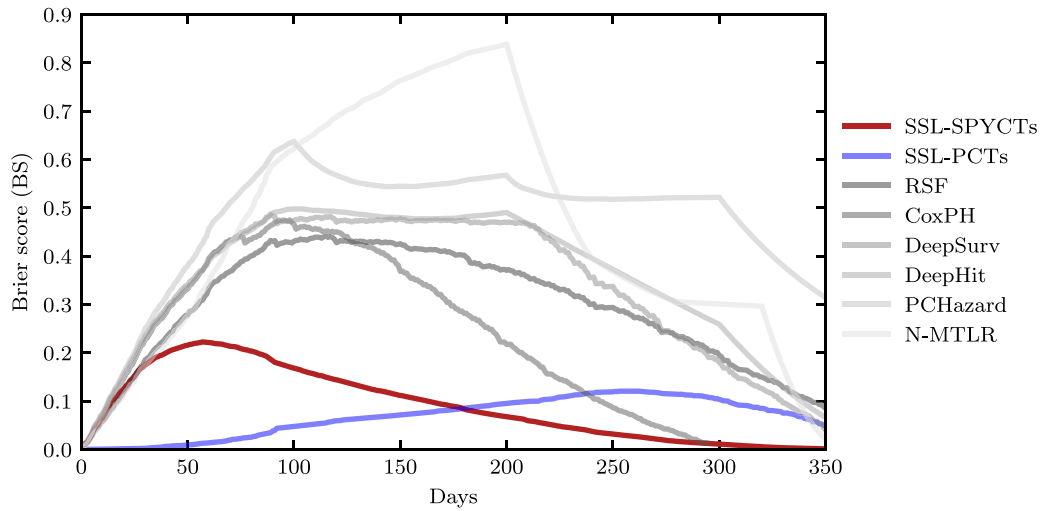
Fig. 4. The performance of different survival analysis models in terms of Brier score for a one-year time interval.

- Semi-supervised Predictive Clustering Trees (PCTs) (Kocev et al., 2013)
- Cox proportional hazards model (Cox, 1972);
- Random survival forests, which is an extension of the Random Forest model specifically developed for survival problems where we have right-censored data. Ishwaran, Kogalur, Blackstone, and Lauer (2008);
- Neural Multi-task logistic regression model, a model that learns a multi-layer perceptron model for the discrete time intervals derived from the time-interval of interest (Yu, Greiner, Lin, & Baracos, 2011);
- DeepSurv, a Cox proportional hazards based on deep neural network (Katzman et al., 2018);
- DeepHit, a model which directly estimates the joint distribution of events in discrete time (Lee et al., 2018); and
- PCHazard, a model which assumes that the continuous-time hazard function is constant in some predefined intervals (Kvamme & Borgan, 2021)

Some of the hyperparameter values of these models were tuned through the grid search method. For the DeepSurv, DeepHit, and PC-Hazard models, the learning rate of the optimiser $\epsilon \in \{0.001, 0.005, 0.01, 0.05, 0.1\}$ was taken into account. For the PCTs and random survival forests, the number of trees in the ensemble $NT \in \{1, 50, 100, 150, 200, 250, 500\}$ and the maximum depth of a tree $d \in \{3, 4, 5, 7, 10\}$ were taken into account.

5.4. Performance comparison

The performance of survival model is most commonly evaluated using the Concordance Index (Brentnall & Cuzick, 2016) and the Brier score BS (Gerds & Schumacher, 2006). One limitation of the Concordance Index is that it does not consider the magnitude of the error in the predicted probabilities. The BS overcomes this limitation by incorporating the magnitude of the error into the calculation (Gerds, Cai, & Schumacher, 2008). Hence we use it as an evaluation metric in our experiments. The BS is used to evaluate the accuracy of a predicted survival function at a given time t; it represents the average squared distances between the observed survival status and the predicted survival probability and is always a number between 0 and 1, with 0 being the best possible value. BS measures the accuracy of the time-to-event predictions and can be considered analogous to a squared error loss. However, if the dataset contains samples that are right censored, then it is necessary to adjust the score by weighting the squared distances

using the inverse probability of censoring weights method (Kvamme & Borgan, 2023). Let $\hat{G}(t) = P(C > t)$ be the estimator of the conditional survival function of the censoring times calculated using the Kaplan–Meier method, where C is the censoring time. Then BS is given by the following expression (Kvamme & Borgan, 2023):

$$BS(t) = \frac{1}{M} \sum_{i=1}^{N} \left(\frac{\hat{S}(t_i, \mathbf{X}_i) \cdot 1_{T_i \geq t, \delta_i = 1}}{\hat{G}(T_i^-)} + \frac{(1 - \hat{S}(t_i, \mathbf{X}_i)) \cdot 1_{T_i > t, \delta_i = 0}}{\hat{G}(T_i)} \right) \quad (15)$$

where M is the number of test instances, $\mathbf{X}_i$ is the feature vector describing the example, $\hat{S}(t, \mathbf{X}_i)$ is the predicted survival probability for time point t_i, δ_i is a boolean value defined by (1), and T_i and T_i^- are the survival times when δ_i is 0 and 1.[1] We calculated BS at each timestamp to see how the performance of the method changes when making predictions further into the future.

To estimate performance, k-fold cross-validation was used, and the value of the parameter k was set to 5 (Hastie, Tibshirani, Friedman, & Friedman, 2009). The predictions are in such a form that we have an estimated survival curve for each person.

6. Results and discussion

The resulting BS curves spanning a period of one year and are shown in Fig. 4. It should be noted that lower BS indicates a more accurate forecast.

Besides just looking at the peak, the Brier score curves provide rich information on the model's performance over the observed period. The semi-supervised SPYCTs outperforms most of the models. One can observe that the PCTs is the only model that achieves lower BS peak than the SPYCTs. However, the Brier curve of the PCTs is right-skewed, while the Brier curve of the SPYCTs is left-skewed, meaning that one model is performing well where the other model is performing relatively badly. The Brier score curve of the PCTs is right-skewed because of a high number of censored instances in the final quarter of the time period, where the model performs significantly worse compared to the previous quarters. It is important to note that a low Brier score does not necessarily indicate that a model is performing well, as other factors, such as bias and overfitting, may also affect the model's performance. Performance analysis of the SPYCT model under various settings and on different datasets can be found in Stepišnik and Kocev (2021). The

---

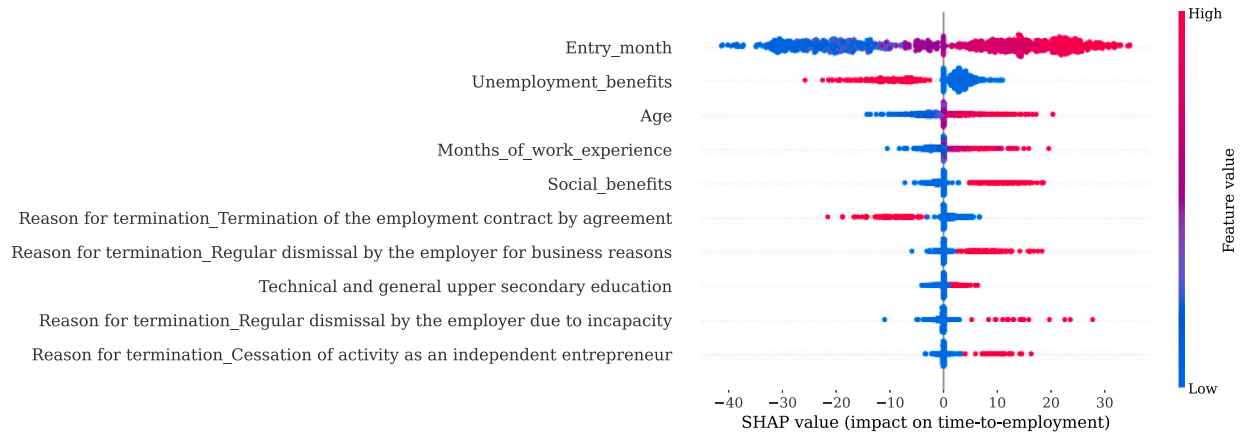[1] The BS metric is implemented in the Python module properscoring.

Fig. 5. SHAP feature importance measured as the mean absolute Shapley values.
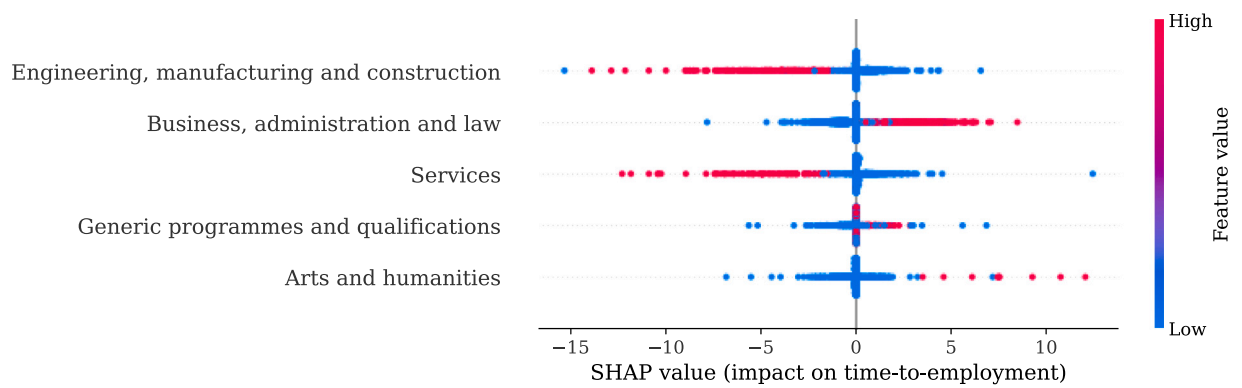

Fig. 6. SHAP feature importance measured as the mean absolute Shapley values for education direction.

performance of the remaining models, aside from SPYCT, on distinct survival analysis datasets is provided in Roy et al. (2022).

6.1. Post-hoc analysis of the feature importances for the model's predictions

Figs. 5–7 show the importances of different features for predicting time-to-employment, derived by using the SHAP method (Lundberg & Lee, 2017). Each instance in the dataset is represented as a dot on each row. The dot's position on the *x*-axis represents the impact of the feature on the model's prediction, indicating whether the particular value leads to a higher or a lower prediction. The dot's colour (from blue to red) conveys information about the values of the features.

The results of the SHAP analysis are shown for three groups of features that turned out to most influence the final PEX:

1. Fig. 5 shows the 6 most important general features,
2. Fig. 6 show the most influential educational directions, and
3. Fig. 7 shows the most influential occupations.

The SHAP analysis in Fig. 5 shows that the most informative feature is the month of entry in the PES. This is certainly the case because of the number of seasonal jobseekers entering the PES in specific periods of the year. Because of the usual high demand and the nature of these job positions, they are expected to be employed soon. This feature is positively correlated with the time-to-employment estimate, which means that certain months of entry significantly affect the outcome of the PES enrolment. The second most informative feature is whether the jobseeker is receiving unemployment benefits. It is negatively correlated with the time-to-employment estimate. For example, the provision of unemployment benefits to temporary jobseekers, such as those in seasonal fields like education and construction, can help them bridge

the gap between jobs and maintain financial stability during periods of seasonal unemployment. This financial support can also provide opportunities for job seekers to acquire new skills and experience that make them more competitive in the job market. Also, receiving unemployment benefits may reduce the stress and anxiety associated with joblessness, which can make it easier for individuals to focus on their job search and perform well in interviews. As a result, those who receive unemployment benefits may be able to secure new employment more quickly than those who do not. On the other hand, receiving social benefits has the opposite effect. Jobseekers tend to wait until the benefits period's expiry before seeking a new job position. This is a well-known effect of social benefits. The third and fourth most informative features are the jobseeker's age and months of work experience, respectively. Their positive correlation with the time-to-employment estimate points to the assumption that people of young age usually get employed sooner than older ones, who have more months of working experience. Finally, for the feature describing the *Reason for termination*, termination by mutual agreement leads to short time-to-employment, i.e., the jobseeker most probably has a plan for future employment. Conversely, termination due to business reasons and dismissal lead to substantially longer time-to-employment.

Fig. 6 shows the SHAP values for the features resulting from the one-hot encoding of the education direction. As expected, engineering and service segments are in demand in the current labour market, leading to shorter time-to-employment. The opposite is true for "business, administration and law" since the labour market is already saturated with candidates from this educational direction.

Finally, Fig. 7 shows the most notable occupations that influence the time-to-employment. As expected, the model captures the fact that health professionals, construction, manufacturing and transportation
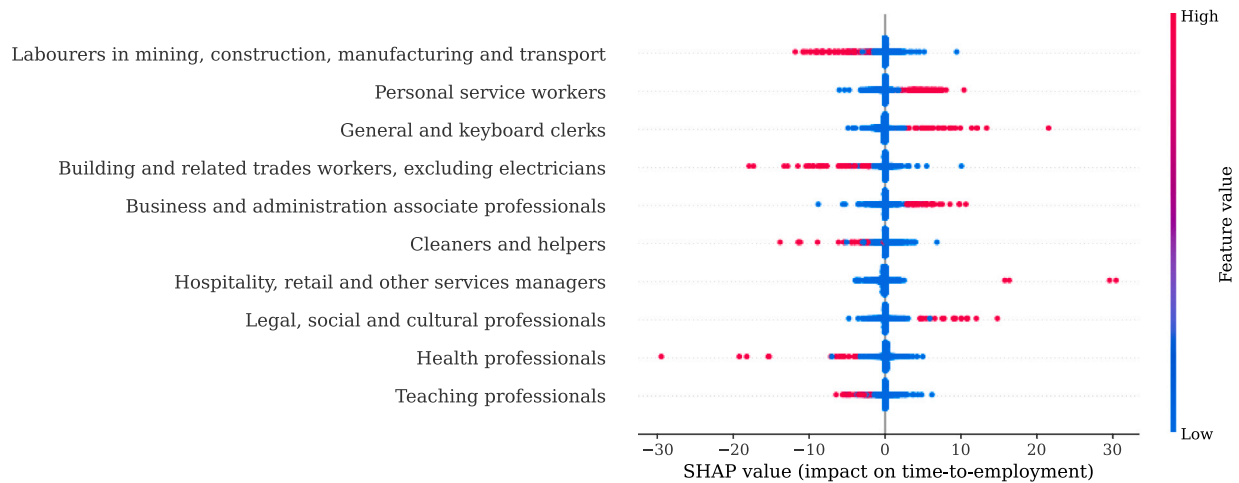
Fig. 7. SHAP feature importance measured as the mean absolute Shapley values — ISCO.

workers, are professions that are in high demand. On the other hand, occupations like legal, social and cultural professionals and general clerks have a longer time-to-employment.

This analysis shows that the model correctly captures the most essential dynamics of the labour market. Apart from time-to-employment predictions, it provides insights into which groups might need additional active labour market policies, as well as strategies for keeping jobseekers from becoming inactive.

6.2. Classification accuracy

Although our main task was not the classification of jobseekers, transforming the predictions of survival analysis models into a classification framework offers another way of quantifying the models performance. Most of the reported results on PEX model performance from different solutions in various PES focus on classification accuracy.

Comparison with other PEX models on BS would have been valuable to provide a more comprehensive analysis. However, due to limited access to the same data and reported models, such comparisons were not feasible. Therefore, the comparison against other PEX models, built from data from other countries, is made in terms of classification accuracy.

For that purpose, the data set is divided into two classes. The first represents those jobseekers that got employed within one year. The second class are those that remained unemployed over the same period. In such a way, the problem is transformed into a problem of binary classification. Cases censored before 180 days were excluded from the dataset due to the unavailability of definitive true labels for them. Cases, that are censored after 180 days were labelled in the second class, since they have already surpassed the threshold regardless what happened in the future.

The survival curves for the two classes are shown in Fig. 8. The curves with thick lines represent the average of all the survival curves within the representative groups, while the blurred curves are individual survival curves. It can be easily seen from the graph that the bulk of unemployed persons have a higher probability $S(t)$ of remaining unemployed over time.

The main feature for the classification is the value of the survival probability $S(t)$ at $t = 180$ days. It was calculated for each individual. The value of 180 days was selected since staying unemployed for such a long time significantly negatively influences jobseekers (Shimer, 2012).

The histograms of $S(t = 180)$ values for jobseekers of both classes (values at the dashed vertical line in Fig. 8) are shown in Fig. 9. The area in the intersection of the two distributions is the area where the errors occur. If the model performs reasonably well, it should assign

lower survival probabilities to instances that belong to the group of jobseekers employed before 180 days and higher probabilities to instances belonging to the group of jobseekers that remained unemployed after 180 days from entering PES. In an ideal situation, the respective survival distributions should not intersect.

Using this result, it is possible to evaluate the model in a classification framework by analysing the ROC curve (Bradley, 1997). The ROC curve values can be easily calculated from the empirical distributions shown in Fig. 9. In that manner, True Positive Rate (TPR) can be estimated from the power of a test, and the False Positive Rate (FPR) is the Type-I error rate (Junge & Dettori, 2018), and these values are calculated for different locations of the threshold value.

The ensemble of SPYCTs model achieved an Area under the ROC curve (AUC) score of 0.78 at the optimal threshold of $S(t = 180 \text{ days}) = 0.17$. This result is obtained by ignoring the cases that were censored before 180 days. To comprehensively assess the model's performance, we conducted two additional simulations, labelling the censored examples before 180 days as either "employed" or "unemployed" separately. The AUC score was 0.76 when all censored examples were labelled as "unemployed" and improved to 0.79 when labelled as "employed". These scores can be viewed as the best and the worst case scenarios. Such an assessment provides insight into the model's performance under different assumptions, enhancing the comprehensive evaluation of its accuracy (see Fig. 10).

The dataset is balanced between the classes. This is easily seen by calculating the accuracy of the trivial model on the dataset. In our case, a trivial model represents a model which categorises every provided jobseeker in the first category, meaning he gets employed in under 180 days. The trivial model achieves 53.4% accuracy, meaning 53.4% of the jobseekers in our test subset are labelled as class 1.

Published evidence on classification accuracy of already implemented solutions is summarised in Table 2. References to the publications and the type of underlying models are also provided in this table.

Even in this case, the performance comparison between those models and our model is not straightforward. Neither the datasets nor the actual implementations of the models listed in Table 2 are publicly available. Therefore, the only possible comparison is through the reported accuracy results. Such a comparison is valid under the assumption of comparable quality and information content in the datasets and ignores the differences between labour markets. In most of the published experiments, the model accuracy of predicting the correct probability of exit over time is close to 70%. The achieved classification accuracy of the ensemble of SPYCTs model is above the average
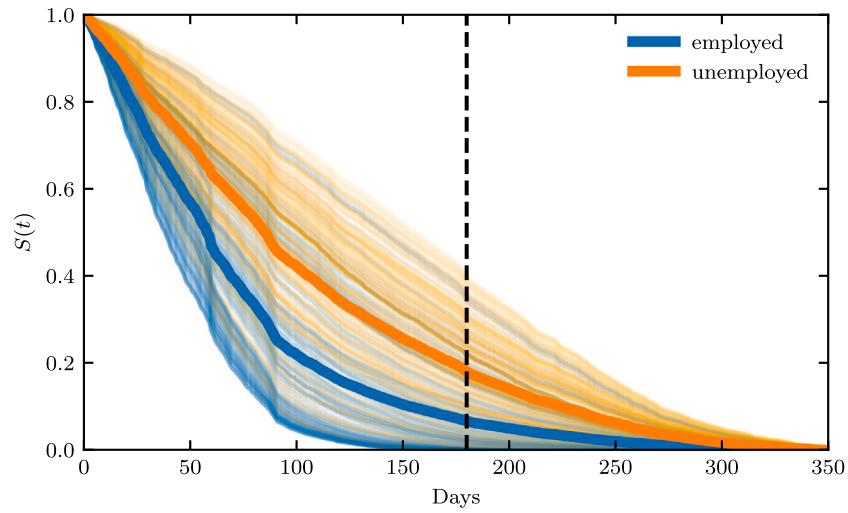
Fig. 8. Survival curves (obtained by the SPYCT model) of jobseekers from two groups — one comprising persons that got employed, and the other persons that remained unemployed during the observed time period. The bold lines are the average of the survival curves across all persons from the respective groups of jobseekers.
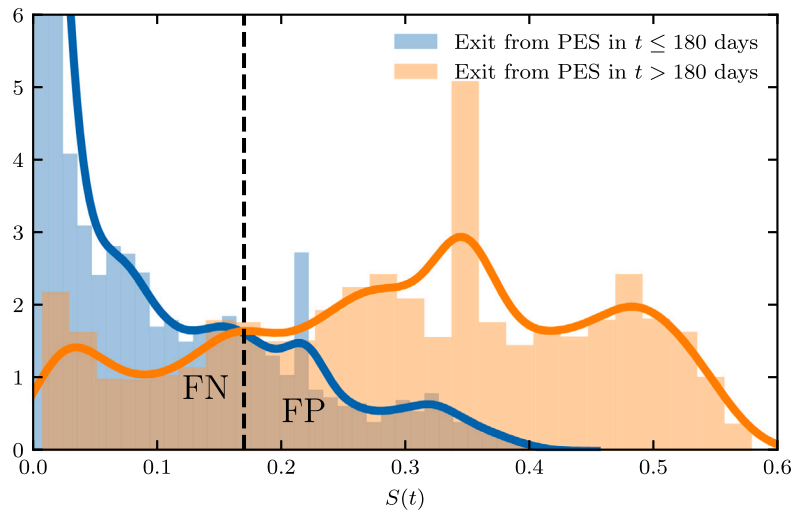

Fig. 9. Probability of exit distributions on a sample of the dataset. One group consists of jobseekers that got employed within 180 days, and the other of jobseekers that are still unemployed after 180 days from the initial entering to PES.
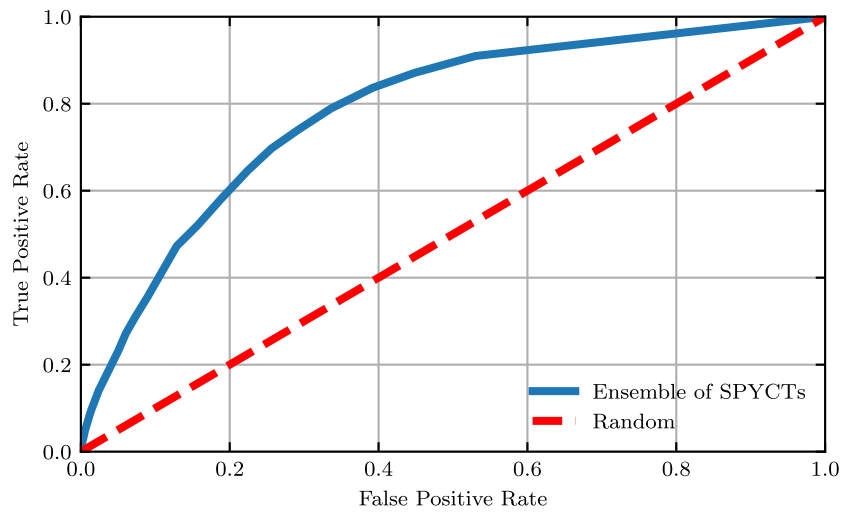
Fig. 10. ROC curve for the ensemble of. SPYCTs model performance in a classification framework. The diagonal dashed line represents the performance of a random classifier.

Table 2

Reported results on classification accuracy by models addressing jobseekers profiling.

Reference	Model	Classification accuracy
Australia (Ponomareva & Sheen, 2013)	Logistic regression	Not reported
Austria (Desiere, Langenbucher, & Struyven, 2019)	Logistic regression	80%–85%
Belgium (Desiere et al., 2019)	Random Forest	67%
Croatia (Pojarski, 2014)	Logistic regression	69%
Finland (Riipinen, 2011)	Statistical model	89%
France (Berthet & Bourgeois, 2014)	Logistic/ Random Forest/ Neural Networks	70%
Ireland (O'Connell et al., 2009)	Probit regression	69%
Netherlands (Wijnhoven & Havinga, 2014)	Logistic regression	70%
New Zealand (Sengul, 2017)	Random forest and Gradient boosting	63%–83%
Slovenia (Boškoski et al., 2021)	Variational Bayes Neural Network	78%

reported in the literature. We, therefore, conclude that the model is well-performing.

7. Conclusion

In this study, we have demonstrated the capabilities of semi-supervised oblique predictive clustering trees (SPYCTs) in a semi-supervised multi-target regression setting for predicting the time-to-employment of a jobseeker. Compared to established methods, ensembles of SPYCTs exhibit better performance as measured through Brier Score (BS) curves. In terms of peak BS, it is competitive to semi-supervised PCTs. However, the main advantage of SPYCTs is in the computational complexity and, thus, scalability to high-dimensional inputs.

The SPYCTs ensemble can be used as a decision support tool in PES to prioritise jobseekers who are in immediate need of PES services, as they will be less likely to get employed soon. It enables the counsellors to focus their attention and resources on the most critical jobseekers with respect to the expected employment duration. With the analysis of the model's predictions, we have identified the most important features of jobseekers that determine the time until their employment. This is another important contribution for PES and jobseekers, giving a clear direction on where they should focus their future efforts.

We have also demonstrated the difference in the survival curves of two specific groups of jobseekers, those who got employed and the other that remained unemployed for one year. By examining the produced survival curves, it is evident that the model distinguishes the two groups very well, and the survival probabilities of representative jobseekers from the unemployed group remain higher over the time period of interest. The model is intended for use in a real-work environment through integration in a web application.

We have demonstrated the interpretability of the model's decisions through the use of the SHAP method, analysing the influence of each feature on survival time. However, a limitation of our approach is its partial explainability due to oblique splits and the use of ensembles, hindering a complete understanding of the decision-making process. To address this, future research could explore modelling the parameters as random variables, aiming to derive optimal probability distributions instead of single optimal values. This potential enhancement may eliminate the need for ensembles, enabling sampling of diverse predictions

from a single-tree model. By examining the probability distributions of parameters within each node and path of the tree, this approach could offer improved explainability of the model's predictions.

As part of the future work, we could also explore the use of the model in other related applications and datasets to further validate its effectiveness. Additionally, further investigation into the model's underlying feature importances and their impact on jobseekers' employability could provide a more in-depth understanding of the model's predictive capabilities.

CRediT authorship contribution statement

Declaration of competing interest

Data availability

The data that has been used is confidential.

Acknowledgements

References

Andonovikj, V., Boškoski, P., & Boshkoska, B. M. (2021). Estimating client's job-search process duration. In *Proc. 24th Intl. Multi-conference Information Society: Slovenian Conference on Artificial Intelligence.*

Andonovikj, V., Boskoski, P., Evkoski, B., Redek, T., & Boshkoska, B. M. (2022). Community analysis in Slovenian labour network 2010–2020. *Journal of Decision Systems*, *31*(sup1), 308–318. http://dx.doi.org/10.1080/12460125.2022.2070944.

Anon (2016). In J. P. Klein, H. C. van Houwelingen, J. G. Ibrahim, & T. H. Scheike (Eds.), *Handbook of survival analysis.* Chapman and Hall/CRC, http://dx.doi.org/10.1201/b16248.

Babińska, M., Chudek, J., Chełmecka, E., Janik, M., Klimek, K., & Owczarek, A. (2015). Limitations of cox proportional hazards analysis in mortality prediction of patients with acute coronary syndrome. *Studies in Logic, Grammar and Rhetoric*, *43*(1), 33–48. http://dx.doi.org/10.1515/slgr-2015-0040.

Ballot, G., Kant, J.-D., & Goudet, O. (2016). Un modèle multi-agent du marché du travail français, outil d'évaluation des politiques de l'emploi : l'exemple du contrat de génération. *Revue Économique*, *Vol. 67*(4), 733–771. http://dx.doi.org/10.3917/reco.674.0733.

Berthet, T., & Bourgeois, C. (2014). Towards 'activation-friendly' integration? Assessing the progress of activation policies in six European countries. *International Journal of Social Welfare*, *23*(S1), http://dx.doi.org/10.1111/ijsw.12088.

Boškoski, P., Perne, M., Rameša, M., & Boshkoska, B. M. (2021). Variational Bayes survival analysis for unemployment modelling. *Knowledge-Based Systems*, *229*, Article 107335. http://dx.doi.org/10.1016/j.knosys.2021.107335.

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7), 1145–1159. http://dx.doi.org/10.1016/s0031-3203(96)00142-2.

Brentnall, A. R., & Cuzick, J. (2016). Use of the concordance index for predictors of censored survival data. *Statistical Methods in Medical Research*, 27(8), 2359–2373. http://dx.doi.org/10.1177/0962280216680245.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 34(2), 187–202. http://dx.doi.org/10.1111/j.2517-6161.1972.tb00899.x.

Desiere, S., Langenbucher, K., & Struyven, L. (2019). Statistical profiling in public employment services: An international comparison. In *OECD social, employment and migration working papers*. Paris: OECD Publishing, http://dx.doi.org/10.1787/b5e5f16e-en.

Fotso, S. (2018). Deep neural networks for survival analysis based on a multi-task framework. http://dx.doi.org/10.48550/arXiv.1801.05512, arXiv preprint arXiv:1801.05512.

Gerds, T. A., Cai, T., & Schumacher, M. (2008). The performance of risk prediction models. *Biometrical Journal*, 50(4), 457–479. http://dx.doi.org/10.1002/bimj.200810443.

Gerds, T. A., & Schumacher, M. (2006). Consistent estimation of the expected brier score in general survival models with right-censored event times. *Biometrical Journal*, 48(6), 1029–1040. http://dx.doi.org/10.1002/bimj.200610301.

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction, Vol. 2*. Springer.

Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. (2008). Random survival forests. *The Annals of Applied Statistics*, 2(3), http://dx.doi.org/10.1214/08-aoas169.

Ittner, A., & Michael, S. (1996). Discovery of relevant new features by generating non-linear decision trees. (pp. 108–113).

Junge, M. R. J., & Dettori, J. R. (2018). ROC solid: Receiver operator characteristic (ROC) curves as a foundation for better diagnostic tests. *Global Spine Journal*, 8(4), 424–429. http://dx.doi.org/10.1177/2192568218778294.

Katzman, J. L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., & Kluger, Y. (2018). DeepSurv: personalized treatment recommender system using a cox proportional hazards deep neural network. *BMC Medical Research Methodology*, 18(1), http://dx.doi.org/10.1186/s12874-018-0482-1.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. http://dx.doi.org/10.48550/arXiv.1412.6980, arXiv preprint arXiv:1412.6980.

Kocev, D., Vens, C., Struyf, J., & Džeroski, S. (2013). Tree ensembles for predicting structured outputs. *Pattern Recognition*, 46(3), 817–833. http://dx.doi.org/10.1016/j.patcog.2012.09.023.

Kretowska, M. (2020). Oblique survival trees in discrete event time analysis. *IEEE Journal of Biomedical and Health Informatics*, 24(1), 247–258. http://dx.doi.org/10.1109/jbhi.2019.2908773.

Kureková, L. M., et al. (2014). Review of profiling systems, categorization of jobseekers and calculation of unit service costs in employment services-implications and applications for slovakia. *Central European Labour Studies Institute (CELSI) Research Reports*, (8).

Kvamme, H., & Borgan, Ø. (2021). Continuous and discrete-time survival prediction with neural networks. *Lifetime Data Analysis*, 27(4), 710–736. http://dx.doi.org/10.1007/s10985-021-09532-6.

Kvamme, H., & Borgan, A. (2023). The brier score under administrative censoring: Problems and a solution. *Journal of Machine Learning Research*, 24(2), 1–26.

Lee, C., Zame, W., Yoon, J., & der Schaar, M. V. (2018). DeepHit: A deep learning approach to survival analysis with competing risks. *AAAI Conference on Artificial Intelligence*, 32(1), http://dx.doi.org/10.1609/aaai.v32i1.11842.

Li, Y., Dong, M., & Kothari, R. (2005). Classifiability-based omnivariate decision trees. *IEEE Transactions on Neural Networks*, 16(6), 1547–1560. http://dx.doi.org/10.1109/tnn.2005.852864.

Li, H. J., Song, S., Tan, W., Huang, Z., Li, X., Xu, W., et al. (2022). Characterizing the fuzzy community structure in link graph via the likelihood optimization. *Neurocomputing*, 512, 482–493. http://dx.doi.org/10.1016/j.neucom.2022.09.013.

Li, H., Xu, W., Qiu, C., & Pei, J. (2023). Fast Markov clustering algorithm based on belief dynamics. *IEEE Transactions on Cybernetics*, 53(6), 3716–3725. http://dx.doi.org/10.1109/tcyb.2022.3141598.

Lundberg, S., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774.

Niklas, J., Sztandar-Sztanderska, K., Szymielewicz, K., Baczko-Dombi, A., & Walkowiak, A. (2015). Profiling the unemployed in Poland: social and political implications of algorithmic decision making. *Fundacja Panoptykon, Warsaw Google Scholar*.

O'Connell, P. J., McGuinness, S., Kelly, E., & Walsh, J. R. (2009). *National profiling of the unemployed in Ireland*. ESRI.

Pereira, G. K., Guilardi, L. F., Dapieve, K. S., Kleverlaan, C. J., Rippe, M. P., & Valandro, L. F. (2018). Mechanical reliability, fatigue strength and survival analysis of new polycrystalline translucent zirconia ceramics for monolithic restorations. *Journal of the Mechanical Behavior of Biomedical Materials*, 85, 57–65. http://dx.doi.org/10.1016/j.jmbbm.2018.05.029.

Pojarski, P. I. (2014). *Malawi-strengthening safety nets systems-MASAF IV: P133620-implementation status results report: Sequence 02: Technical Report*, The World Bank.

Ponomareva, N., & Sheen, J. (2013). Australian labor market dynamics across the ages. *Economic Modelling*, 35, 453–463. http://dx.doi.org/10.1016/j.econmod.2013.07.038.

Riipinen, T. (2011). Risk profiling of long-term unemployment in Finland. In *Power point presentation at the European commission's "PES to PES dialogue dissemination conference," Brussels, September* (pp. 8–9).

Roy, B., Stepišnik, T., Vens, C., & Džeroski, S. (2022). Survival analysis with semi-supervised predictive clustering trees. *Computers in Biology and Medicine*, 141, Article 105001. http://dx.doi.org/10.1016/j.compbiomed.2021.105001.

Sengul, G. (2017). Learning about match quality: Information flows and labor market outcomes. *Labour Economics*, 46, 118–130. http://dx.doi.org/10.1016/j.labeco.2017.04.001.

Shimer, R. (2012). Reassessing the ins and outs of unemployment. *Review of Economic Dynamics*, 15(2), 127–148. http://dx.doi.org/10.1016/j.red.2012.02.001.

Stepišnik, T., & Kocev, D. (2020). Oblique predictive clustering trees. http://dx.doi.org/10.1016/j.knosys.2021.107228.

Stepišnik, T., & Kocev, D. (2021). Semi-supervised oblique predictive clustering trees. *PeerJ Computer Science*, 7, Article e506. http://dx.doi.org/10.7717/peerj-cs.506.

Wang, P., Li, Y., & Reddy, C. K. (2017). Machine learning for survival analysis: A survey. http://dx.doi.org/10.48550/ARXIV.1708.04649.

Wijnhoven, M. A., & Havinga, H. (2014). The work profiler: A digital instrument for selection and diagnosis of the unemployed. *Local Economy*, 29(6–7), 740–749. http://dx.doi.org/10.1177/0269094214545045.

Yang, J. S., Nam, H. J., Seo, M., Han, S. K., Choi, Y., Nam, H. G., et al. (2011). OASIS: Online application for the survival analysis of lifespan assays performed in aging research. In E. Scalas (Ed.), *PLoS ONE*, 6(8), Article e23525. http://dx.doi.org/10.1371/journal.pone.0023525.

Yizhong, X., Lin, Z., Baranchenko, Y., Lau, C. K., Yukhanaev, A., & Lu, H. (2017). Employability and job search behavior. *ER*, 39(2), 223–239. http://dx.doi.org/10.1108/er-02-2016-0042.

Yu, C. N., Greiner, R., Lin, H. C., & Baracos, V. (2011). Learning patient-specific cancer survival distributions as a sequence of dependent regressors. *Advances in Neural Information Processing Systems*, 24.

Zhang, C., Du, Z., Yang, Y., Gan, W., & Yu, P. S. (2021). On-shelf utility mining of sequence data. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 16(2), 1–31.