The Survival Analysis of Mental Fatigue Utilizing the Estimator of Kaplan-Meier and Nelson-Aalen

R. Eswar Reddy and K. Santhi(✉)

School of Computer Science and Engineering, Vellore Institute of Technology, Vellore 632014, Tamil Nadu, India
regantieswar.reddy2021@vitstudent.ac.in, santhikrishnan@vit.ac.in

Abstract. The aim of this study is to investigate mental fatigue using the Kaplan-Meier and Nelson-Aalen estimators in survival analysis. Mental fatigue is a common occurrence when the mind becomes tired from regular tasks, and it can have a negative impact on an employee's operational functions and job efficiency. To detect mental fatigue, the shallow Kaplan-Meier method is employed by analyzing data from employee burnout evaluations.

Both the Kaplan-Meier and Nelson-Aalen estimators have proven to be effective in automatically analyzing various features from raw data. However, they often impose a significant burden on system resources during training and predictions. Therefore, alternative methods of analysis are necessary to derive the survival curve.

In this paper, we provide a mathematical foundation for the Kaplan-Meier method and explain the concept of censoring, including right censoring, interval censoring, and left censoring. Furthermore, we construct a Kaplan-Meier survival curve, which represents the probability of survival over time. The Kaplan-Meier survival curve is considered the most reliable and is recommended for predicting the variable under investigation, particularly in the fields of public health and medical research.

The findings of this research can also be utilized to develop interventions and strategies aimed at reducing mental fatigue and improving employee morale. Enhancing employee morale can positively impact an organization as a whole, as mental fatigue has been associated with lower job satisfaction and an increased likelihood of employee turnover, both of which can be further explored in future studies. Overall, this study sheds light on the significance of understanding and addressing mental fatigue in the workplace, and it provides valuable insights that can contribute to the well-being of employees and the success of organizations.

Keywords: mental fatigue · Kaplan Meier · Nelson-Aalen · fatigue survival curve · burnout

P. Pareek et al. (Eds.): IC4S 2023, LNICST 536, pp. 218–241, 2024.
https://doi.org/10.1007/978-3-031-48888-7_19

1 Introduction

Employees who report feeling good about themselves and their working conditions tend to be more productive overall. As a result, they contribute to the success of the business or organization. Nevertheless, the situation in the majority of businesses has changed as a result of the pandemic. Almost 69% of the workforce has been experiencing burnout since implementing work-from-home and office policies. The percentage of employees who have burned out is high. There has been a rise in the number of businesses caring about their workers' emotional well-being. This trend can be attributed to the growing recognition of the negative impact that burnout can have on employee productivity and overall business success. As a result, many companies are implementing programs and policies aimed at preventing burnout and promoting mental health in the workplace. To counteract this, we plan to develop a web application that businesses may use to track staff burnout. Additionally, employees themselves can use it as a tool to monitor burnout and evaluate mental health in the hectic workplace.

This study also found that the amount of mental fatigue [1] is related to how much pain, anxiety, and depression affect how much fatigue [2] affects a person's life, and these things work together in a cycle that can make fatigue worse and keep it going. There are many possibilities for the effects of each of the others. 80% of employees with depression also report sleeping poorly, supporting the idea that the two are related. There is a strong correlation between this condition and poor sleep quality, with decreased sleep efficiency being the primary sleep issue described. Mental fatigue makes it worthwhile to investigate this K-M method. According to the literature and our most recent findings, women have a higher prevalence of mental fatigue than men. Women's work is handed to us, working women are handed the kids, the homework, kids, jobs, etc.

Instead of investigating employees in the organization to check for mental fatigue disorder [3], the organization checks the health condition of employees and mentally how much struggle they are facing, effective working on ongoing work, whether the organization has given additional responsibility or reduced project delivery on time, reviews meetings and client meetings in the cases of team members, makes plans to change the upcoming one, earns new things, and talks to family. Employees' work burden stress, and keeping on top of the status in the daily meeting, do not support the work environment.

Life is moving at a breakneck speed, and the demands of work and school are only getting heavier. Extended mental exertion of any kind necessitating undivided attention inevitably results in tiredness, with all the unpleasant consequences that entail slower reaction times, dizziness, nausea, etc. Hence, in order to aid in reducing its harmful consequences, it is important to recognize [4] and study various forms of weariness.

The role of emotion in human interaction, understanding, and decision-making has grown in recent years. Recognizing others' feelings is a fundamental skill for establishing rapport in everyday life. Emotional state is assessed using a person's EEG signal [5], which measures the amount of brain activity during various conditions. To date, the best method for extracting human emotions from EEG [6] data has been a 6 layer feed forward neural network that has been subjected to extensive biological testing. Functionalities including preprocessing, feature extraction, and classification were all provided by this system. First, a band pass filter is used to do EEG preprocessing [4,7]. Several methods of electrophysiological recording have been used to shed light on the intricate brain interactions that underlie cognitive tiredness in situations of mental fatigue, allowing researchers to better understand the neural circuits at play in this state [8].

This percentage is inversely related to the number of people with fatigue. A decreasing curve starting at 1 represents the Kaplan-Meier estimates of survival time graphically [9]. The size of the steps depends on factors such as his length of residency, the likelihood of a mental fatigue disorder, and the possibility that he will suffer through the allotted time without ever encountering the event of interest. Censored observations refer to data that is missing either temporarily or permanently, and they can occur at any point in the research paper. There are essentially three distinct categories of censorship. The most prevalent type of censorship is called right censorship, and it occurs when a patient is observed for a certain amount of time without experiencing the event of interest. Hence, there is a gap in the survival time [10] series on the right side of the observational period. We know that the event of interest does not occur for this patient until the censoring date, the second type of censorship occurs when the event of interest occurs between two unknown dates and we do not know which date it occurred on.

The zero Reynolds number and long wavelength assumptions, It is discovered that, for a given flow rate, the pressure rise lowers as the peripheral layer viscosity drops, and that, for a given non-zero pressure drop, the flow rate increases as the viscosity of the peripheral layer decreases [11]. The beneficial in creating a physically suitable workplace and promoting professional productivity. Furthermore, they should consider permitted working hours based on people's abilities, reducing tension factors, creating a culturally and ethically secure atmosphere, respecting new ideas, providing spiritual and mental well-being, planning, reconstructing the system of instruction, utilizing recommendations and criticism, and other factors. These factors can prevent emotional exhaustion within the organization. Since depersonalizing has a direct correlation with performance, the managers can strengthen coworker relationships by implementing counselling programmers [12] and communication skills training. The electrodes on an EEG device collect electrical impulses that communicate at different EEG frequencies. The Fast Fourier Transform [13] is a method that can be used to identify these raw EEG signals as discrete waves with a variety of frequencies. A group of learners who perform poorly on the learning curve come together to form an ensemble of classifiers [13,14], also known as a committee of classifiers. Learning a large number of less effective classifiers and integrating them in a certain manner is the purpose of the ensemble of classifiers technique rather than learning a single, efficient classifier. Data mining algorithms aid in the analysis and prediction of large data sets with minimal human intervention [15].

Predicting and analyzing diabetes can be done with a number of data mining programs. Fast and accurate automated algorithms for summarizing text and generating a summary that can be spoken aloud. An automated decision-making system's variance could be reduced, which would improve the system's accuracy. In the intervening time, ensemble systems have been utilized effectively to solve a wide range of machine learning challenges, including feature selection, confidence estimation, missing features, incremental learning, error correction, class-imbalanced data, and learning concept drift from non stationary distributions, just to identify a few. By doing in-depth research on technological aid, user experience, and health care, we can help reduce the deadly risks that people face and be ready to act quickly in emergencies [16]. Because of its usefulness in so many programs, neural model-based text analysis has recently gained traction. Researchers have identified and justified a large number of techniques for enhancing text analytics effectiveness. Text categorization, text generation, text summarizing, query formulation, query resolution, and sentiment analysis are just some of the areas where these methods have been put to good use [17]. To achieve uniform scale L2 regularizes of linear models may assume that all features are centered on zero or have variance in the same order. Because these things are often used in the objective function of learning algorithms. This method of presentation has advanced greatly since the days of rainbow-hued spreadsheets. With the advent of datasets [18] another execution time and space complexity of mining has drastically decreased. Retrieval techniques went to a whole new level after [19] receiving this data. If all commercial activities cease, the business will cease to exist. Operating alone, the application server could not sustain startup costs [20]. Statistical methods for regression and classification are incorporated into machine learning algorithms. Sensors are used to capture data, which is then transmitted to the Blynk app. The automatic water controller only activates when the relative humidity falls below a predetermined threshold [21]. To accompany the connected device, we need a resource provisioning system that is easily managed; this is only feasible if we have accomplished cloud service models [22]. Every plant has specific requirements that must be met to ensure its survival. Therefore, it is necessary to establish a system where plants can communicate with the user [21]. The proposed strategy uses a technique for reducing the number of dimensions and clustering similar objects together. For both symmetrical and asymmetrical data sets, it provides the highest accuracy for the larger of the two [23].

Third, left censoring, occurs when someone from a certain fatigue is known to have the event before a certain date, but the time period between the occurrence of the event and the specific date is unknown. The zero Reynolds number and long wavelength assumptions, It is discovered that, for a given flow rate, the pressure rise lowers as the peripheral layer viscosity drops, and that, for a given non-zero pressure drop, the flow rate increases as the viscosity of the peripheral layer decreases [11]. Survival probability utilizing the Kaplan Meier (KM) survival estimator, Nelson-Aalen estimator, and Hazard Model based on regression could be used to assess prediction scores for Mental fatigue [24]. In this paper, an

Employee burnout datasets was used to (i) analyses the research to identify key differential mental fatigue, (ii) examine survival associated with most altered mental fatigue using the web-based Kaplan Meier and Nelson-Aalen Plotter tool, and (iii) evaluate the possibility of the potential at the datasets between mental fatigue and control variables. During the construction of survival time probabilities and curves, the serial duration's for specific participants are ordered from shortest to longest regardless of when they entered the research. By employing this technique, all subjects within the group commence the analysis at the same point and are all surviving until mental fatigue persons are identified. Two outcomes are possible: 1) the subject can see the event of interest, or 2) they may be censored. This subject's total survival time cannot be determined precisely due to censorship. This can occur when an adverse occurrence for the research occurs, such as the Employee dropping out, being lost to follow-up, or required data not being available, or when something positive occurs, such as the research getting before the subject observed the event of interest, they survived at least until their conclusion of the research, but it is unidentified [25] what occurred to them afterward. Thus, censorship can occur either during the research. Metal fatigue is a risk factor that causes some of their diseases, like Constantly feeling overwhelmed or stressed, Cynicism, uncertainty, and pessimism Depression, anxiety, and suicidal thoughts Sleep disruptions and pattern alterations, Tension, pain, and headaches Digestion problems and recurring colds High blood pressure, abnormal heart rate, brain fog, and strokes Obesity and cardiovascular disease.

2 Method

Dataset. Employee burnout is a dataset collected from the kaggle website [26]. The following are the data attributes and their descriptions, we are implement the python code [27]. The shape of data frame is: (20633, 9) instances will be there in dataset. Data preprocessing the data is processing is onehotEncoder is technique will represent the text, category variable is numeric values, the transform the text and category variables data is the will transform to numeric values. Data is pre processing the data is find out the missing and noise data, duplicate data. The data is replace the noise data and duplicate data by the using the means values, the replace the missing values by the means values on the dataset.

Employee ID. The distinctive ID that the company gives to each employee.

Date of Joining. The day the individual began working for the company.

Gender. In the box plot, we see that some of the women's burnout rate data points are significantly different from the rest. This is something that needs to be handled by us. Female workers are over represented in the data. It shown the Fig. 2.

The gender of the employee. Men have a higher burnout rate than women do on average. Let's investigate why this would be the case by looking at how the two sexes fare in other areas, such as titles and hours worked. It shown the Fig. 1.

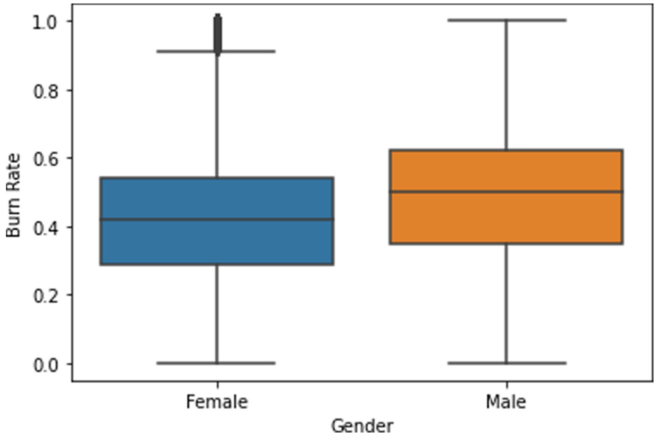

Fig. 1. The gender difference in burn rate

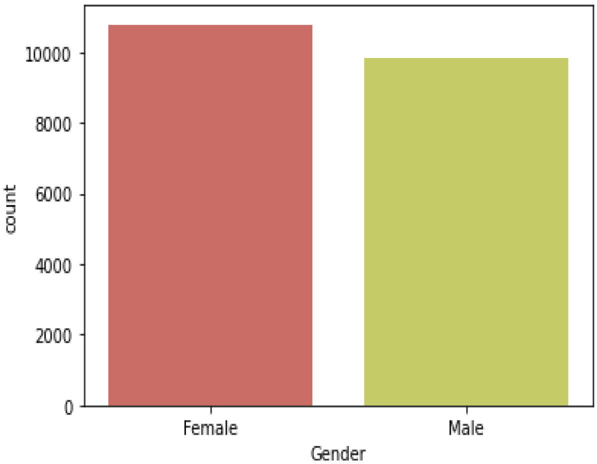

Fig. 2. The gender and count of employees

Designation. The employee's position in his or her organization. In the interval [0.0, 5.0], "0.0" is the least significant digit, and "5.0" is the most significant. More men than women hold positions with a designation of 2.0 or above. It shown the Fig. 3.

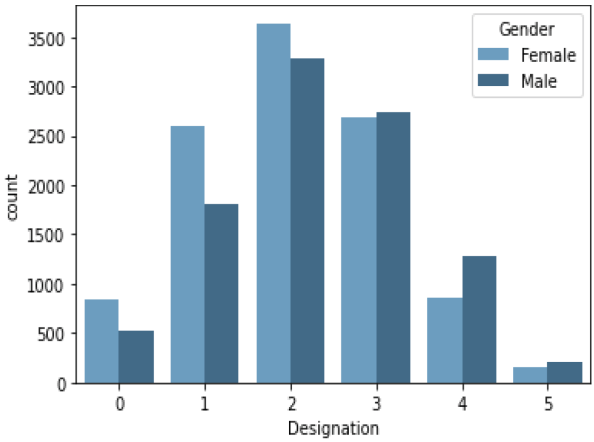

Fig. 3. Designation for men

Resource Allocation. The number of resources given to an employee for work, which should be thought of as the number of working hours. Between one and ten higher means more resources. Most women work up to 8 h a day, whereas most men work up to 10 h. The median number of hours worked by men and women differs by one hour. It shown the Fig. 4.

Company Type. Employee The figure of resource allocation for an employee-based company displays how the available resources are distributed among the employees to achieve the company's goals. It helps in identifying which employee has been assigned what task and how much time and resources have been allocated to it. May classify their employers based on the services or products they offer. It shown the Fig. 5.

WFH Setup Available. Is the worker allowed to work from his or her home office.

Mental Fatigue Score. A number from 0 to 10 that shows how mentally tired the worker is at work, where 0 means no mental fatigue and 10 means extreme mental fatigue.

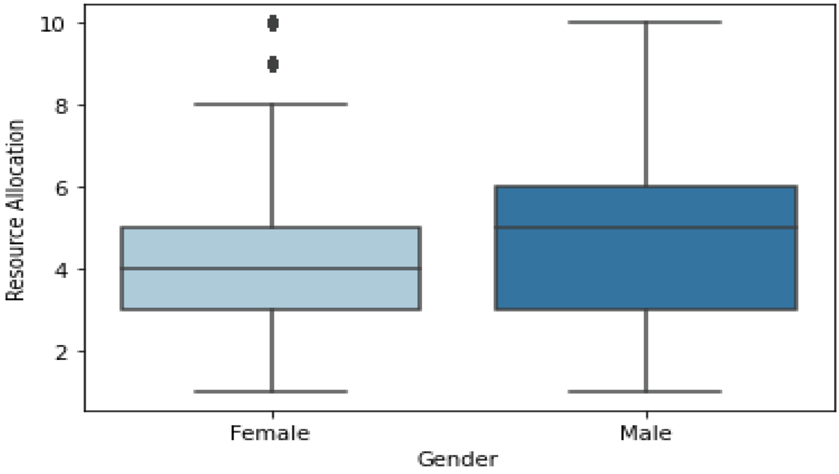

Fig. 4. Resource Allocation

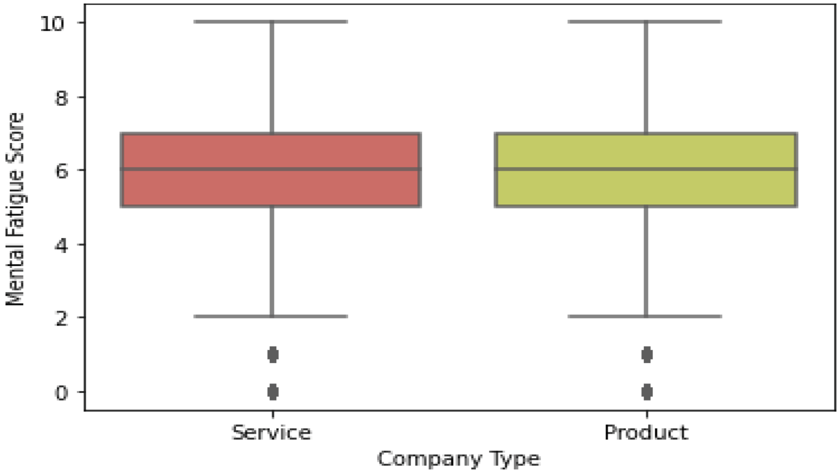

Fig. 5. Company Type

Burn Rate. The target value in each worker's data shows the rate of burnout while on the job. Values from 0.0 to 1.0 show that burnout is getting worse. The correlation between fatigue score and burn rate appears to be very significant. It is important to address burnout in the workplace, as it can have negative impacts on both employees and organizations. Employers should consider implementing strategies to prevent and manage burnout, such as promoting work-life balance and providing support resources. It shown the Fig. 6.

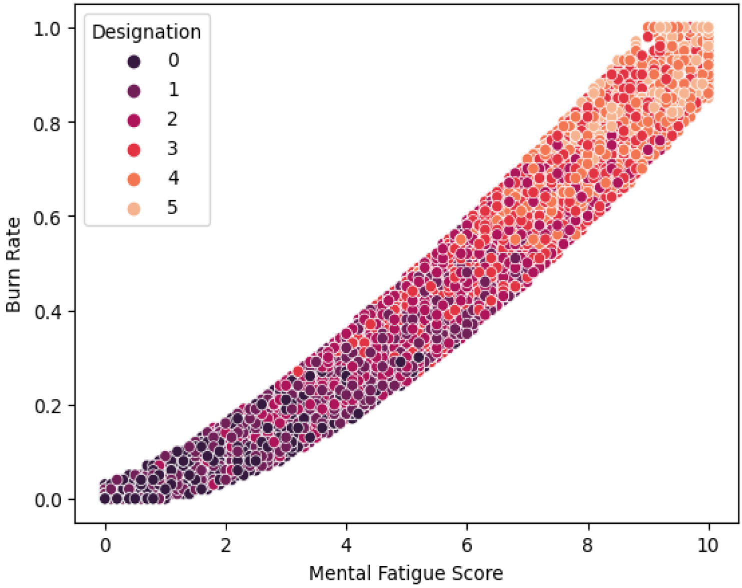

Fig. 6. Mental fatigue and burnout

The Kaplan-Meier Estimator. The Kaplan-Meier curves and survival predictions, there is now a better way to look at data when your mind is tired. The Kaplan-Meier estimator [28] is often used to describe how long a population will fatigue because it gives estimates of the survival function that are not based on statistics. If you want to know how likely a patient is to live past a given point after receiving mental health treatment, one of the most reliable statistical techniques is the KM estimate. When working with incomplete or hidden data, the KM method is especially helpful because it lets all available data points be used in the analysis. Additionally, it can be used to compare survival rates between different treatment groups or populations. This method of making graphs and charts is very user-friendly. The effectiveness of an intervention in mental fatigue research is determined by tracking how many people are rescued or made healthy thanks to the intervention. Notwithstanding the complexity of some subjects or settings, KM estimation is the simplest method for predicting longevity over time. Events, censorship, and the likelihood of survival can all be estimated with the help of the Kaplan-Meier curves.

The Kaplan-Meier survival curve [29] is a useful tool in statistics for analyzing time-to-event data and making comparisons between groups of people. The number of people who avoid dying during a specific time period can be calculated using the survival curve. This can be analyzed for an evaluation of a pair of patient populations or subjects as well as the statistical difference in their overall survival.

Depending on how much time has passed, either the product limit estimator or the Kaplan-Meier curve can be used to figure out the survival function. If the data are already organization into intervals, if the sample size is large, or if a large population is of interest, a clinical life table analysis may be more time-efficient. Both of these approaches will be addressed in greater detail. The Kaplan-Meier survival curve is the cumulative chance of survival over many intervals of time. This analysis relies on three presumptions. To proceed, we believe that censored employees at any given time have the same survival and growth as unregulated employees; second, we think that those who participate in research at the beginning or the end will have significantly different survival times. Furthermore, we'll presume that the event occurs just when it's expected to. In situations where the occurrence would pick up the energy level and bring routine work, this can be done from the office or home. This approach can be particularly useful for individuals who have flexible work arrangements or who are able to work remotely. It allows them to balance their personal and professional responsibilities while still being productive.

If an employee is followed up on more often and for shorter periods of time, it is possible to predict how long they will stay with the company. This can be beneficial for employers in terms of retention strategies and succession planning. However, it is important to balance this with the need for employee autonomy and trust in the workplace. The term "product limit estimate" can be used to describe the Kaplan-Meier estimate. The method involves calculating the odds of an occurrence happening at a given instant. To arrive at a final estimate, we multiply these probabilities by any previously calculated probabilities. If you want to know your probability of surviving at any given moment in time, just plug those numbers into the following Eq. 1.

The Kaplan-Meier method or Kaplan-Meier curve can be made with just two pieces of information: the time until the event of interest and the status of the patient at that moment. In medical research, the Kaplan-Meier method is often used to figure out how likely it is that a patient with a certain disease will live. It is a non-parametric statistic that takes into account time-to-event data and censored observations. Let $D1 < D2 < ... < Dn$, $i < N$ be a collection of separate ordered times finding the mental fatigue times observed in N individuals; in a given time Di (i = 1, 2,..., n), the number $di \geq 5$ of mental fatigue are observed, and the number ri of subjects, whose either mental fatigue or censored time is greater than or equal to Ti, are deemed "at risk", the observed times to event. T In its simplest form, there is no need to elaborate on the formula for the conditional probability of survival past time Ti, which is simply that the Kaplan-Meier method is an estimate of the conditional probability of survival at different periods in time identified by the event's likelihood of occurring. The Kaplan-Meier method is commonly used in medical research to estimate the survival rate of patients with a particular disease. It takes into account the occurrence of events such as death or relapse to calculate the probability of survival at different time intervals.

$$P(D_i) = \frac{r_i - d_i}{r_i} \tag{1}$$

By dividing the total number of subjects by the total number of patients at each point in time, one can calculate the likelihood of survival. Subjects who get lost because they are tired are not counted as having a disorder. Instead, they are considered "censored" and taken out of the denominator. By multiplying the individual survival probabilities at each interval leading up to that point by the law of multiplication of probabilities, one can determine the cumulative probability of survival up to that point. This approach is commonly used in medical research to account for the fact that some subjects may drop out of a study for reasons unrelated to the disorder being studied. By censoring these subjects, researchers can more accurately estimate the probability of survival for those who remain in the study.

The term conditional probability describes this type of probability. Because of the limited number of events, the calculated probability at any given interval is not very precise. However, the overall probability of survival at each point is Estimating the survival function at time Δt is done by multiplying the conditional probability of survival at that time by the formula.

The conditional probability of survival ($\bar{S}(\Delta t)$), also called cumulative probability or cumulative survival, is the chance that a patient will be mentally tired days after enrolling in a study, if the patient has been alive for at least Δt days before enrolling. In a hypothetical situation where a patient in an intensive care unit checks his or her level of fatigue and lives for hours per day, the product rule of conditional probabilities says that the cumulative survival is the product of survival probabilities. This information is useful in medical research because it assists clinicians and researchers. A patient's conditional probability of survival, also called cumulative probability or cumulative survival, is the chance that he or she will be mentally tired days after enrolling in a study, given that the patient has been alive for at least t days before that. Understand the long-term effects of treatments and interventions on patients' survival and quality of life. Additionally, it can aid in making informed decisions about patient care and treatment plans.

$$\bar{S}(\Delta t) = \prod_{i:D_i<t} P(D_i) = \prod_{i:D_i<\Delta t} (1 - \frac{d_i}{r_i}) \tag{2}$$

The definition of its variance is Variance is a statistical measure that quantifies the amount of variability or dispersion in a set of data. It is calculated by taking the average of the squared differences from the mean of the data set.

$$\sigma(\bar{S}(\Delta t)) = \bar{S}(\Delta t)^2 \sum_{i:D_i<\Delta t} \frac{d_i}{r_i(r_i - d_i)} \tag{3}$$

Variance is an important tool in statistical analysis, as it helps to understand the spread of data points around the mean and can be used to make predictions about future data. But outliers and extreme values in the data set might affect it, and it might be necessary to fix them before using the variance to draw any conclusions. Because of the censoring, r_i is not simply equal to the difference between $r_i - 1$ and $d_i - 1$, the right approach to calculate r_i is $r_i = r_{i-1}d_{i-1}C_{i-1}$, where $C_i - 1$ is the number of censored cases between $D_i - 1$ and D_i. This calculation is commonly used in statistics to determine the variability of a data set. However, when dealing with censored data, a modified approach must be taken to accurately calculate the variability.

3 Result and Discussion

The Kaplan-Meier method [30] is a deft statistical analysis of survival times that not only provides for filtered observations in the right way but also makes use of the information from filtered individuals up to the point of filtering. While investigating the effects of mental fatigue, it is typical to use two interventions and evaluate the outcome in terms of the employee's ability to stay fatigued. Hence, the Kaplan-Meier approach [31]is a valuable resource that may have an essential role in producing evidence-based data on expected survival [32]. Survival analysis is a type of statistical analysis used to examine an incident that happened relatively frequently over a particular period of time. Hence, it seeks to discover how often something occurs. The word "survival" can mean anything from the employee's depression, mental tiredness, mental illness, sleeping, and anything in between. Populations, or collections of recipients who are monitored throughout time, are used in survival studies to record significant medical events as they occur and associate them with an intervention of interest. Survival analysis requires the determination of the "survival time", [10] which is the amount of time that has passed since the baseline date and before the event happens. As a result, it is essential to determine whether the employees in question witnessed the incident of interest or were prevented from doing so by the filters. For the statistical application to determine the cumulative probability of the event, it is essential to know this information. The Kaplan-Meier method's [2,33] primary function is to generate survival curves as a function of time, providing a visual representation of the clinical phenomenon being investigated. The ordinate of a Kaplan-Meier curve shows the cumulative survival time, while the abscissa shows the elapsed time. The time intervals used to create a Kaplan-Meier curve aren't decided upon in advance but rather are determined by the occurrence of events. The Kaplan-Meier technique is preferable because it estimates for correction.

Plotting Survival Curves Using matplotlib.pyplot, Seaborn, Kaplan-MeierFitter, CoxPHFitter
The Kaplan-Meier [34] curves for survival time are unappealing to the eye. Enhanced plots can be created using matplotlib.pyplot, seaborn, KaplanMeier-Fitter, CoxPHFitter libraries [27]. The following sections show and describe Kaplan-Meier curves generated with matplotlib.pyplot, seaborn [35], Kaplan-MeierFitter [30], and CoxPHFitter. These libraries' functions are using the graphs generated [36] by the KaplanMeierFitter graphs [31], which find the Kaplan Meier estimates, the x-axis is the mental fatigue score, the y-axis is the designation of the employees, find out the fatigue score of the employees. The Kaplan-Meier curves are useful for analysing survival data and estimating the probability of an event occurring over time. They are commonly used in medical research to analyse patient outcomes and can also be applied to other fields such as finance and engineering. It shown the Fig. 7 in the graph.

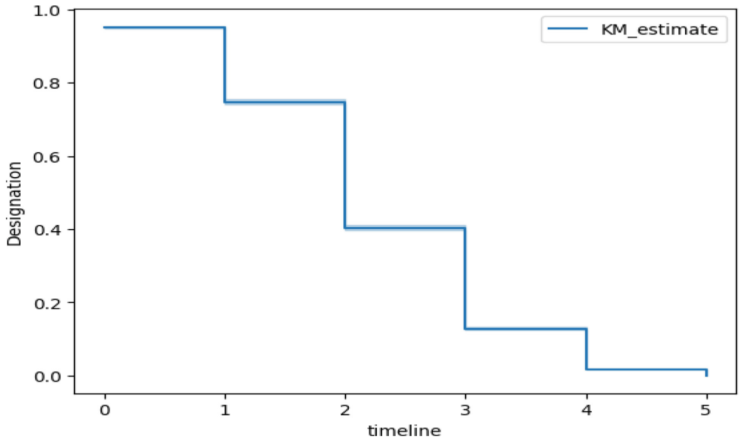

Fig. 7. Kaplan Meier estimate curve

The median survival time is 2.0, and the designation is 0.5. This suggests that half of the patients survived beyond 2.0 units of time and the other half did not, and the designation of 0.5 indicates that the survival probability at 2.0 units of time is approximately 50%.

The Survival Function. Survival, denoted by $(\bar{S}(\Delta t))$, is the probability that T happens before Δt, where Δt is any moment during the observation. In survival analysis, the survival function is often used to figure out how likely it is that something will happen at a certain on time, like a machine breaking down or an employees fatigue. It is also used to compare survival rates between different groups or treatments. Specifically, the probability that an employee will still be fatigued after a certain burnout of time, denoted by t, is the survival function. It shown the Fig. 8, and Table 1.

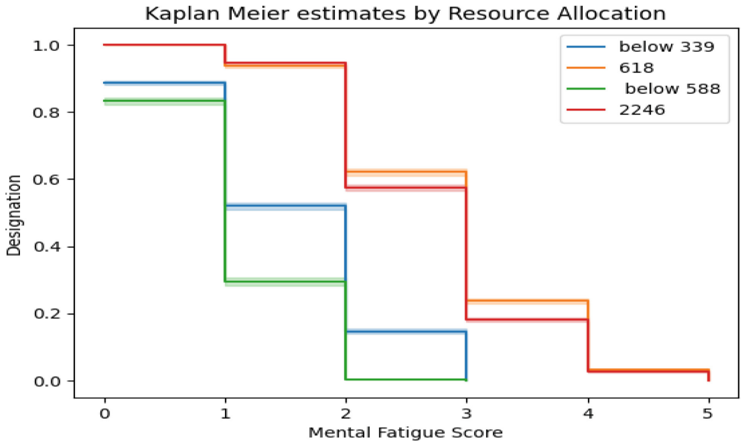

Fig. 8. Survival of different gender groups

Table 1. Kaplan Meier estimate

timeline	Kaplan Meier estimate
0	0.951038
1	0.747703
2	0.404027
3	0.128106
4	0.018177
5	0
Text (0, 0.5, "designation")	

The colour in this figure indicates which clinics correspond to certain curves. Confidence intervals for each time point and overall are shown as a band of shading. At any given period, the plus signs denote the censored instances. The confidence intervals represent the range of values within which the true population parameter is likely to fall. The censored instances refer to observations that are incomplete or truncated, usually due to limitations in data collection or follow-up. Mental fatigue has an increased survival curve, therefore, more patients remain there than in fatigue. It is advised that research be conducted into the reasons so many fatigued employees end up leaving. It's unclear if the discrepancy can be attributed to fatigue itself or if the employees were chosen for fatigue, depression, mental illness, mental stress, mental tensions, or some other factor. The research could also explore potential solutions to address the high turnover rate, such as implementing flexible work schedules, providing mental health resources, or offering additional support for employees experiencing fatigue. It is important for employers to prioritise the well-being of their employees in order to maintain a productive and healthy work environment. It shown the Fig. 9.

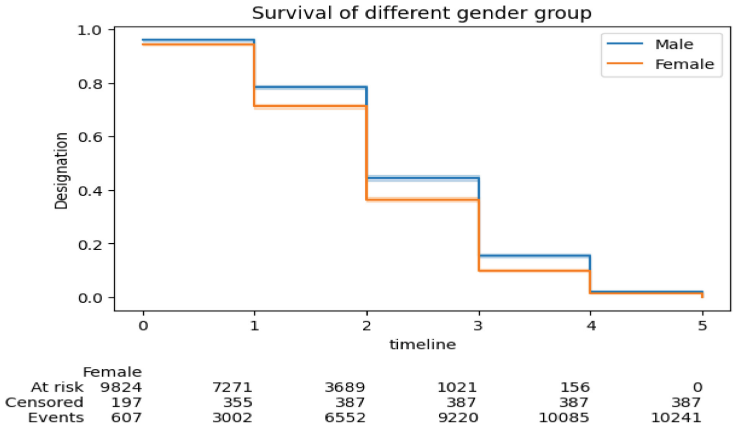

Fig. 9. Survival function

To calculate the survival probability, we'll use the Kaplan-Meier estimator [29,39], which relies on a function called the survival function $-(\bar{S}(\Delta t))$. The curve depicts how the probability of survival changes over time. Persons suffering from mental fatigue have a lower chance of survival. The comparable graph can be created with the 95% confidence interval. A failed curve can also be drawn. It is the inverse of survival, probability, and cumulative density. The median survival time and 95% confidence intervals are estimated next. This can be accomplished with the median survival time and median survival times() functions. In this case, the median survival time is 2.0 h, implying that 50% of the sample lives for 2.0 h and 50% experiences mental fatigue during that period. The 95% KM estimate lower time is 2.0, while the KM estimate upper time is 2.0. We can examine the difference involving discrete categories using the KM estimate. However, this method is only applicable when the variable has fewer categories. A mask filtering object where males are true, and a plotting object. The Plot curves for Male and Female observations. The curve demonstrates that the aggregate survival probabilities of female patients are higher than those of male patients at any given time Fig. 9.

The survival function provides the probability that the event has occurred within the time interval t. The cumulative density is the complement of the

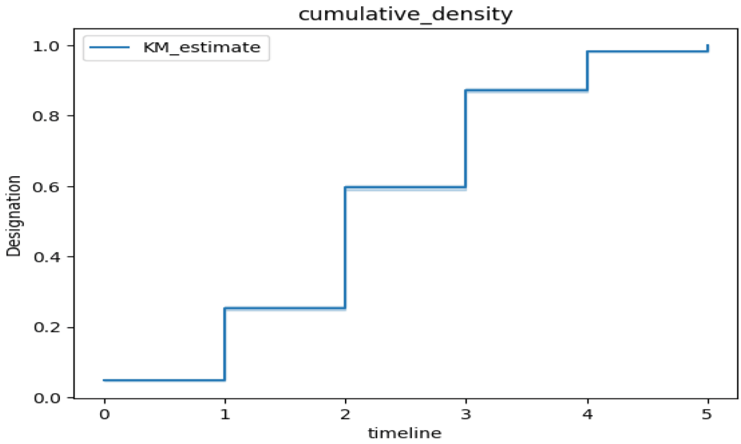

Fig. 10. Cumulative density in km estimator

survival function, which is the probability that the event has taken place by time t. Reading our plots, we can see using cdf that there is a 50% chance that the event has transpired within 5 h. This is a range, as indicated by the blue lines' broad intervals. This complement, the survival function, indicates that there is a 60% chance that the event has not occurred within 5 h.

The survival probability is the chance that a worker won't get tired between an expected point in time and a future point in time. To demonstrate, if (9944) = 0.9, the employee survival probability shrinks to 0.1. These statistics will be eliminated if the employee survives the completion of the investigation. This paper's Kaplan-Meier estimator is 0.9, achieving the best result and predicting the method However, it is important to note that the Kaplan-Meier estimator is not a foolproof method and may have limitations in certain situations. Therefore, it is crucial to consider other factors and data points before making any final decisions based solely on this statistic.

Estimating Hazard Rates Using Nelson-Aalen

The survival function is an important way to describe and show how well the model works. There's an additional method, though. Unfortunately, the Kaplan-Meier [28,37] estimate is sometimes transformed to yield information on the population hazard function (t). This transformation is known as the Nelson-Aalen estimator [37,38], which estimates the cumulative hazard function. It is a non-parametric method used in survival analysis to estimate the hazard rate from lifetime data. The Nelson-Aalen hazard function is used for this function. The Nelson-Aalen hazard function is a non-parametric estimator that is particularly useful when the hazard rate changes over time. It provides an estimate of the cumulative hazard function, which can be used to estimate the population hazard function. The Nelson-Aalen hazard function is a non-parametric estimator of the cumulative hazard function that makes no assumptions about the underlying distribution of survival times. It is particularly useful when analysing data with complex censoring patterns.

$$\bar{H}(\Delta t) = \sum_{D_i \leq \Delta t} \frac{d_i}{r_i} \tag{4}$$

where di is the number of fatigued employees at time t and ri is the number of employees at the start. Survival functions are where basic survival analysis starts, but cumulative hazards are where more complicated methods begin. It Show the Table 2 and Fig. 10.

Table 2. hazard function

timeline	NA estimate
0	0.058806
1	0.338227
2	1.008063
3	2.292282
4	4.168251
5	5.3682
<AxesSubplot:xlabel='timeline'>	

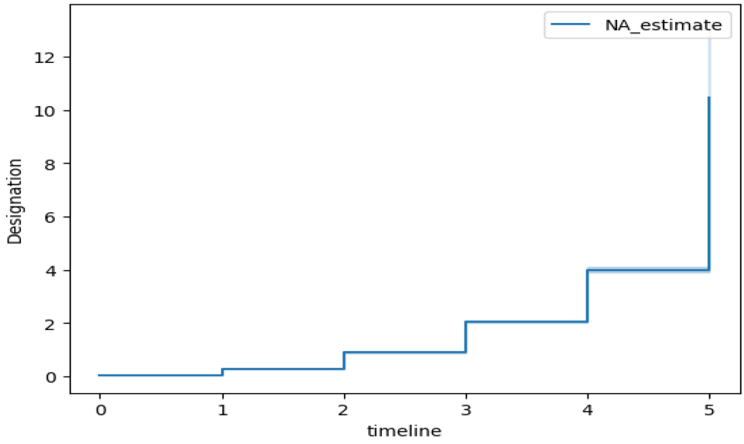

Fig. 11. Hazard function curve in NA estimate

This research predicts values at a certain point 10, If the hazard probability, denoted by $\bar{H}(\Delta t)$, is true at a given time, an employee under observation has event fatigue at that time. If the value of (9944) is equal to 0.9, for instance, the probability that the employee is still mentally fatigued is predicted, and the employee's fatigue levels are checked to examine the burnout. The hazard function, in contrast to the survival function, is the occurrence of an event. Having a lower hazard probability and a higher survival probability [33] is good for the employee. The hazard function can be used to predict the likelihood of an event occurring in the future, such as an employee leaving a company. By using the hazard function, employers can take proactive measures to reduce turnover and improve retention rates. The hazard function is a useful tool for predicting the likelihood of an event occurring in a given time frame. Employers can identify potential risks and take steps to mitigate them by analysing the hazard function, resulting in a safer work environment for their employees. In the paper we analysed to find the best result and predict the method are would use, we investigated the employees to check the mental fatigue and burnout status in organisations (Fig. 12).

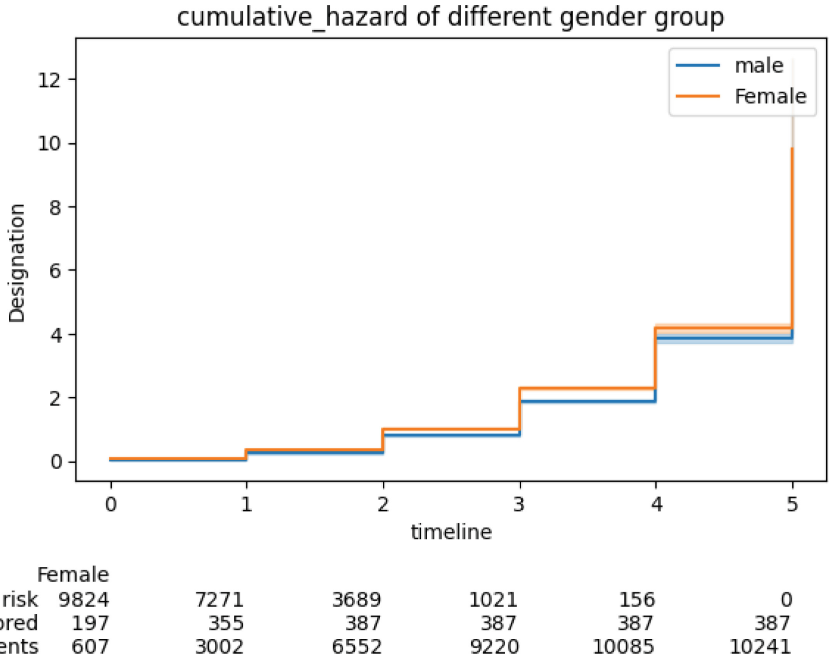

Fig. 12. Cumulative hazard function curve in NA estimate

To calculate the survival probability, we'll use the Nelson-Aalen estimator. The survival probability is the chance that a worker won't get tired between an expected point in time and a future point in time. To demonstrate, if (10373) = 0.04, the employee survival probability shrinks to 10.36. These statistics will be eliminated if the employee survives the completion of the investigation. It shown the Fig. 11. This paper's Nelson-Aalen is 0.9, achieving the best result and predicting the method However, it is important to note that Nelson-Aalen estimator is not a foolproof method and may have limitations in certain situations. Therefore, it is crucial to consider other factors and data points before making any final decisions based solely on this statistic. The estimate the number of persons will be at the high risk will be 100 employee is risk in the affect the mental fatigue due to the designating, the estimated the number of persons will be at the fatigue persons on below the 200 persons affect the fatigue at risk.

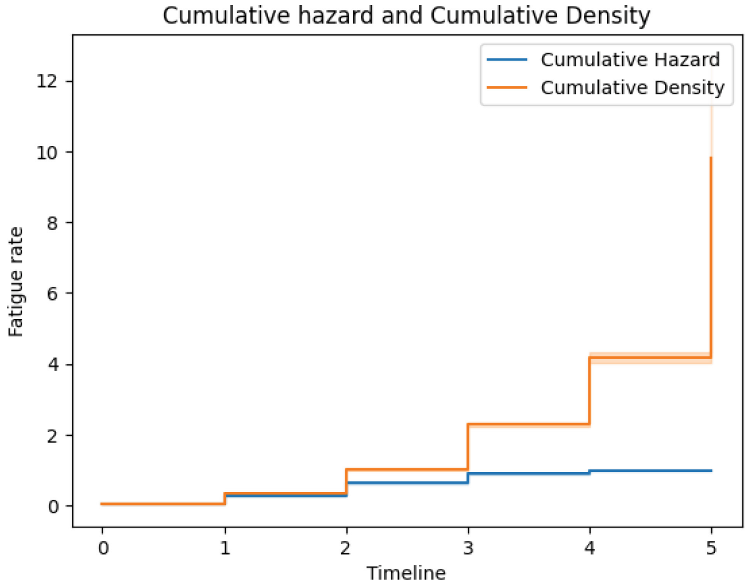

Fig. 13. Cumulative hazard function curves

The survival probability of the divided int two group that is male and female, estimated and predicting the employee work and meta fatigue, estimated the male's will be faster fatigue, more female is fatigue is more and compare t the male check the status abut the graph Fig. 13, the Cumulative density function using find the estimated the km estimate and Cumulative hazard function using estimated the predicting the employees affect the meta fatigue, the point the function the risk, It assists to consider actual phenomena and the way their hazard functions may be shaped. If T represents the designation of an employee when it develops fatigue for the first time, then one might expect the corresponding hazard function h(t) to increase with time; that is, the conditional probability of a serious fatigue in the time will increase with the employee's designation and responsibility. In contrast, one might expect h(t) to decrease during the period of responsibility if unpredictable fatigue was being studied in a situation where the employee had a designation and a responsibility. This is known to be the result of selection during time's fatigue. When T is the time at which metal fatigue will have a greater impact on employees due to the w designation and less employee responsibility, the hazard function will remain relatively constant in t. Given that fatigue has not yet occurred, the probability of designation and responsibility in the next time interval does not change with t, but the probability of fatigue in the next designation will increase as the fatigue level rises. Survival analysis relies heavily on the exponential distribution, which is uniquely characterised by this property. The hazard function may take on a more intricate form. If T denotes the mental fatigue of fatal outcome, then the hazard function h(t) is anticipated to decrease initially before progressively increasing at the end, reflecting a higher risk of unpredictable fatigue and fatal outcome.

TTF estimation without the need to know the failure times of all observed units. This would reduce the number of necessary calculations and, more importantly, facilitate the procedure for obtaining data throughout the entire observation period. The optimal method to the NA estimator is to divide the observation period into intervals and evaluate TTF for the limits of these intervals, as opposed to calculating TTF for each failure. This is the Mean time to failure TTF: [0.01665967 0.02371429 0.02838923 ... 0.03362179 0.02897202 0.03997889], mean time to failure TTF: 167 h on the estimated. The predict the estimated at the work environment persons is affect the fatigue at 45 h mean time failures. The TTF estimated the 204.86 value on the NAE values estimated.

The cumulative hazard function is the bathtub curve, which represents the fatigue's life cycle. Combining the hazard rate and the slope of the bathtub curve produces the curve's hazard rate. It shown the Fig. 14.

While the image above shows the hazard rate, the Nelson-Aalen estimator's curve illustrates how the hazard rate varies over time. The concave shape of the cumulative hazard function indicates that we are dealing with a "fatigue fatal outcomes" category of event, where the failure rate is highest early on and decreases over time blue line in the image. On the other hand, the convex shape of the cumulative hazard function indicates that we are dealing with an event red line indicative of unpredictable fatigue. In this paper the Nelson-Aalen estimator of the cumulative hazard function. The estimation of the cumulative hazard function and an intuitive understanding of the results' interpretation. While the Nelson-Aalen estimator is considerably popular than the Kaplan-Meier survival curves, it is still widely used. The ability to provide more precise survival estimates and identify survival differences between subgroups of patients that KM cannot detect, as demonstrated here for various age groups [31]. In this paper The ability to provide more precise survival estimates and identify survival differences between subgroups of patients that KM detect [40], as demonstrated here for various groups.

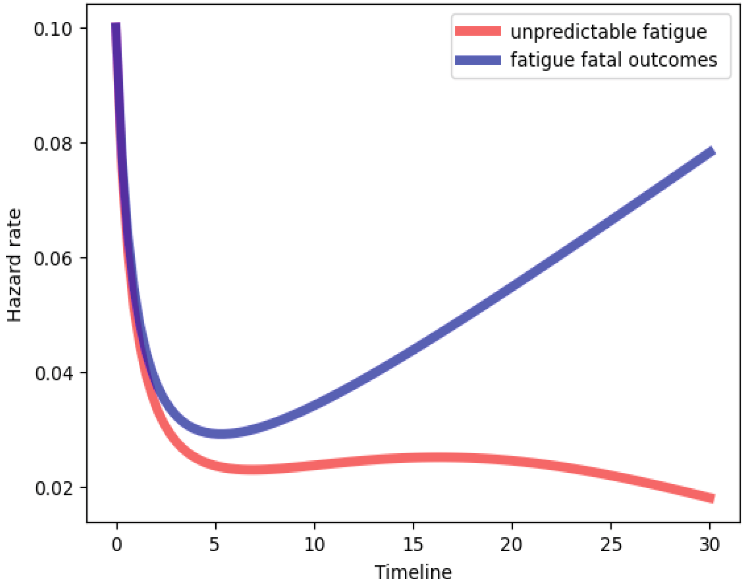

Fig. 14. Time-Dependent Hazard Rate Analysis

4 Conclusion

Mental fatigue is linked to anxiety, depression, and burnout, even though men are less likely to be anxious or sad at the time they are diagnosed. To enhance employees' quality of life and direct the creation of individualized multi component plans according to gender variations, Kaplan Meier's model is trending towards emphasizing psychopathology and early referral to mental health experts. The necessity for experts to establish public policies and provide proper care to better the daily lives of those suffering from fatigue, and the significance of fatigue to the employee's health system as a whole. The article, the effect of Kaplan Meier on the estimated model was seen, and all variables that the method found to be important were used in the regression coefficients. This highlights the importance of utilizing statistical methods to identify significant variables that can contribute to the development of effective policies and care plans for individuals suffering from fatigue. It also highlights the need for continued research and analysis in this area to improve overall employee health and well-being. Using the coefficient of determination, the Kaplan-Meier method is superior to the proposed model, the age of the employees when the fatigue was discovered is crucial and plays a major role in the survival time of employees with mental fatigue, the incidence of fatigue is less important than the gender variable. While analyzing processes connected to survival time is related, it is essential to use the Kaplan-Meier method to assess the significance of the independent factors that can play a significant role in identifying the explanatory variables. This method is the best way to get the analysis and predict the employee's status, but the complexity of this condition necessitates further research to refine the parameters for its treatment.

References

1. Darnai, G., et al.: The neural correlates of mental fatigue and reward processing: a task-based fMRI study. Neuroimage **265**, 119812 (2023). https://doi.org/10.1016/j.neuroimage.2022.119812
2. Sauch Valmaña, G., Miró Catalina, Q., Carrasco-Querol, N., Vidal-Alaball, J.: Gender, mental health and socioeconomic differences in fibromyalgia: a retrospective cohort study using real-world data from Catalonia. Healthcare **11**(4), 530 (2023). https://doi.org/10.3390/healthcare11040530
3. Zhang, Y., Guo, H., Zhou, Y., Xu, C., Liao, Y.: Recognising drivers mental fatigue based on EEG multi-dimensional feature selection and fusion. Biomed. Signal Process. Control **79**, 104237 (2023). https://doi.org/10.1016/j.bspc.2022.104237
4. Xu, X., Tang, J., Xu, T., Lin, M.: Mental fatigue degree recognition based on relative band power and fuzzy entropy of EEG. Int. J. Environ. Res. Public Health **20**(2), 1447 (2023). https://doi.org/10.3390/ijerph20021447
5. Sreeshakthy, M., Preethi, J.: Classification of emotion from EEG using hybrid radial basis function networks with elitist PSO. In: 2015 IEEE 9th International Conference on Intelligent Systems and Control (ISCO), pp. 1–4 (2015). https://doi.org/10.1109/ISCO.2015.7282340
6. Sreeshakthy, M., Preethi, J., Dhilipan, A.: A survey on emotion classification from EEG signal using various techniques and performance analysis. Int. J. Inf. Technol. Comput. Sci. **8**(12), 19–26 (2016). https://doi.org/10.5815/ijitcs.2016.12.03
7. Preethi, J., Sowmiya, S.: Emotion recognition from EEG signal using ISO-FLANN with firefly algorithm. In: 2016 International Conference on Communication and Signal Processing, pp. 1932–1936 (2016). https://doi.org/10.1109/ICCSP.2016.7754508
8. Zorzos, I., Kakkos, I., Miloulis, S.T., Anastasiou, A., Ventouras, E.M., Matsopoulos, G.K.: Applying neural networks with time-frequency features for the detection of mental fatigue. Appl. Sci. **13**(3), 1512 (2023). https://doi.org/10.3390/app13031512
9. D'Arrigo, G., Leonardis, D., Abd ElHafeez, S., Fusaro, M., Tripepi, G., Roumeliotis, S.: Methods to analyse time-to-event data: the Kaplan-Meier survival curve. Oxid. Med. Cell. Longev. **2021**, 1–7 (2021). https://doi.org/10.1155/2021/2290120
10. Vale-Silva, L.A., Rohr, K.: Long-term cancer survival prediction using multimodal deep learning. Sci. Rep. **11**(1), 13505 (2021). https://doi.org/10.1038/s41598-021-92799-4
11. J. Xidian Univ.: Literature review on MHD Peristaltic Transport of non-Newtonian fluids through channels/Tubes, vol. 14, no. 5 (2020). https://doi.org/10.37896/jxu14.5/263
12. Gorji, M.: The effect of job burnout dimension on employees performance. Int. J. Soc. Sci. Hum. **1**(4) (2011). https://doi.org/10.7763/IJSSH.2011.V1.43
13. Anbarasi, M., Durai, M.A.S.: Prediction of protein folding kinetics states using hybrid brainstorm optimization. Int. J. Comput. Appl. **42**(7), 635–643 (2018). https://doi.org/10.1080/1206212x.2018.1479348
14. Anbarasi, M., Durai, M.S.: Incipient knowledge in protein folding kinetics states prophecy using deep neural network-based ensemble classifier. Int. J. Comput. Aided Eng. Technol. **13**(3), 341 (2020). https://doi.org/10.1504/ijcaet.2020.109519

15. Durai, M.A.S., Anbarasi, M., Handa, J.: Prediction of cancer disease using classification techniques in map reduce programming model. In: Advances in Human and Social Aspects of Technology, pp. 139–158. IGI Global (2018). https://doi.org/10.4018/978-1-5225-2863-0.ch007
16. Chellatamilan, T., Kumar, N.S., Valarmathi, B.: Effective deployment of multi-cloud customizable chatbot application for COVID-19 datasets. In: Nagarajan, R., Raj, P., Thirunavukarasu, R. (eds.) Operationalizing Multi-Cloud Environments. EICC, pp. 361–379. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-74402-1_20
17. Chellatamilan, T., Valarmathi, B., Santhi, K.: Research trends on deep transformation neural models for text analysis in NLP applications. Int. J. Recent Technol. Eng. IJRTE **9**(2), 750–758 (2020). https://doi.org/10.35940/ijrte.b3838.079220
18. Santhi, K., Valarmathi, B., Chellatamilan, T.: Depth impurity pruned strategies for extracting high utility itemsets. Int. J. Eng. Technol. **7**(3.4), 52 (2018). https://doi.org/10.14419/ijet.v7i3.4.16747
19. Santhi, K., Chellatamilan, T., Valarmathi, B.: PFBtree for big data memory management system. Indian J. Public Health Res. Dev. **9**(6), 531 (2018). https://doi.org/10.5958/0976-5506.2018.00666.6
20. Suzana, S., Shanmugam, S., Uma Devi, K.R., Swarna Latha, P.N., Michael, J.S.: Spoligotyping of Mycobacterium tuberculosis isolates at a tertiary care hospital in India. Trop. Med. Int. Health **22**(6), 703–707 (2017). https://doi.org/10.1111/tmi.12875
21. Shibani, K., Sendhil Kumar, K.S., Siva Shanmugam, G.: An effective approach for plant monitoring, classification and prediction using IoT and machine learning. In: Dash, S.S., Lakshmi, C., Das, S., Panigrahi, B.K. (eds.) Artificial Intelligence and Evolutionary Computations in Engineering Systems. AISC, vol. 1056, pp. 143–154. Springer, Singapore (2020). https://doi.org/10.1007/978-981-15-0199-9_13
22. Santhi, K., Valarmathi, B., Chellatamilan, T.: Multi-cloud path planning of unmanned aerial vehicles with multi-criteria decision making: a literature review. In: Nagarajan, R., Raj, P., Thirunavukarasu, R. (eds.) Operationalizing Multi-Cloud Environments. EICC, pp. 31–63. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-74402-1_3
23. Nivetha, S., Valarmathi, B., Santhi, K., Chellatamilan, T.: Detection of type 2 diabetes using clustering methods – balanced and imbalanced pima Indian extended dataset. In: Pandian, A.P., Palanisamy, R., Ntalianis, K. (eds.) ICCBI 2019. LNDECT, vol. 49, pp. 610–619. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-43192-1_69
24. Karim, S., et al.: Gene expression study of breast cancer using Welch Satterthwaite t-test, Kaplan-Meier estimator plot and Huber loss robust regression model. J King Saud Univ - Sci **35**(1), 102447 (2023). https://doi.org/10.1016/j.jksus.2022.102447
25. Rich, J.T., Neely, J.G., Paniello, R.C., Voelker, C.C.J., Nussenbaum, B., Wang, E.W.: A practical guide to understanding Kaplan-Meier curves. Otolaryngol. Neck Surg. **143**(3), 331–336 (2010). https://doi.org/10.1016/j.otohns.2010.05.007
26. Are Your Employees Burning Out. https://www.kaggle.com/datasets/blurredmachine/are-your-employees-burning-out
27. Shukl, P.: A complete guide to survival analysis in Python. Aspiring Mach. Learn. Eng. https://www.kdnuggets.com/2020/07/complete-guide-survival-analysis-python-part1.html
28. Burneo, J.G., Villanueva, V., Knowlton, R.C., Faught, R.E., Kuzniecky, R.I.: Kaplan-Meier analysis on seizure outcome after epilepsy surgery: do gender and

race influence it. Seizure **17**(4), 314–319 (2008). https://doi.org/10.1016/j.seizure.2007.10.002
29. Etikan, I.: The Kaplan Meier estimate in survival analysis. Biometrics Biostat. Int. J. **5**(2) (2017). https://doi.org/10.15406/bbij.2017.05.00128
30. Kishore, J., Goel, M., Khanna, P.: Understanding survival analysis: Kaplan-Meier estimate. Int. J. Ayurveda Res. **1**(4), 274 (2010). https://doi.org/10.4103/0974-7788.76794
31. Kiessling, J., Brunnberg, A., Holte, G., Eldrup, N., Sörelius, K.: Artificial intelligence outperforms Kaplan-Meier analyses estimating survival after elective treatment of abdominal aortic aneurysms. Eur. J. Vasc. Endovasc. Surg. (2023). https://doi.org/10.1016/j.ejvs.2023.01.028
32. Elamin, A.M.K., Mohmmed, A.O.A.: The Cox regression and Kaplan-Meier for time-to-event of survival data patients with renal failure. World J. Adv. Eng. Technol. Sci. **8**(1), 097–109 (2023). https://doi.org/10.30574/wjaets.2023.8.1.0183
33. Li, C., Gao, Y., Lu, C., Guo, M.: Identification of potential biomarkers for colorectal cancer by clinical database analysis and Kaplan-Meier curves analysis. Medicine (Baltimore) **102**(6), e32877 (2023). https://doi.org/10.1097/MD.0000000000032877
34. Chandan, Charu: Healthcare Data Analytics. Chapman Hall CRC (2015)
35. scikit-. Introduction to Survival Analysis with scikit-survival. scikit-survival. https://scikit-survival.readthedocs.io/en/stable/userguide/00-introduction.html
36. Stel, V.S., Dekker, F.W., Tripepi, G., Zoccali, C., Jager, K.J.: Survival analysis I: the Kaplan-Meier method. Nephron Clin. Pract. **119**(1), c83–c88 (2011). https://doi.org/10.1159/000324758
37. Carnero Contentti, E., et al.: Neuromyelitis optica spectrum disorders with and without associated autoimmune diseases. Neurol. Sci. (2023). https://doi.org/10.1007/s10072-023-06611-4
38. Elhardt, C., Schweikert, R., Kamnig, R., Vounotrypidis, E., Wolf, A., Wertheimer, C.M.: Recurrence of perforation and overall patient survival after penetrating keratoplasty versus amniotic membrane transplantation in corneal perforation. **261**(7), 1933–1940 (2023). https://doi.org/10.1007/s00417-022-05914-0
39. Lacny, S., et al.: Kaplan-Meier survival analysis overestimates the risk of revision arthroplasty: a meta-analysis. Clin. Orthop. Relat. Res. **473**(11), 3431–3442 (2015). https://doi.org/10.1007/s11999-015-4235-8
40. Ismiguzel, I.: Hands-on Survival Analysis with Python. https://towardsdatascience.com/hands-on-survival-analysis-with-python-270fa1e6fb41