The Influence of the COVID-19 pandemic on the US Freelance Workforce

Name: Lisa de Ruiter
Student Number: u260767

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN DATA SCIENCE & SOCIETY
DEPARTMENT OF COGNITIVE SCIENCE & ARTIFICIAL INTELLIGENCE
SCHOOL OF HUMANITIES AND DIGITAL SCIENCES
TILBURG UNIVERSITY

Thesis committee

Supervisor: dr. Elisabeth Huis in 't Veld
Second reader: dr. Silvy Collin

Tilburg University
School of Humanities & Digital Sciences
Department of Cognitive Science & Artificial Intelligence
Tilburg, The Netherlands
21 January 2022

Table of Contents

Declaration

By submitting this work, I declare that this work is entirely my own, except those parts duly identified and referenced in my submission. It complies with any specified word limits and the requirements and regulations detailed in the assessment instructions and any other relevant programme and module documentation.

In submitting this work, I acknowledge that I have read and understood the regulations and code regarding academic misconduct, including that relating to plagiarism, as specified in the Programme Handbook. I also acknowledge that this work will be subject to a variety of checks for academic misconduct.

Work on this thesis involved the collected data from human participants selected to represent the US workforce. This data was collected by the Edelman Intelligence institute combined with collected data from the Bureau of Labor Statistics. I declare that the original owner of the data and code used in this thesis retains ownership of the dataset during and after the completion of this thesis.

I acknowledge that I do not have any legal claim to this dataset. I have evaluated this project according to the "Ethics checklist Student research with human participants". The dataset used in this thesis is not publicly available and is licensed under The Upwork Inc. platform located in San Francisco, California, United States. The author of this thesis tried not to focus too much attention towards the theoretical conclusions of the report as that would create an unintentional bias view during the conduction of this thesis.

Lisa de Ruiter

Preface

In the past decade, the freelance workforce has increased in the United States into an established type of employment (Atkenson,2020). Online freelancing work or services reflects a form of labor that relies on mediated and fully digital interactions between these workers and potential employers. Freelance service opportunities provide technological and human solutions by bundling organizational, technological and human capabilities and competences (Akhmetshin, 2018). Organizations often lack resources and competences and securing freelancers is crucial for an organization's growth strategy. Focusing on the freelance workforce secured online also means looking at online labor platforms which are the online environments where freelancers find their work. Online labor platforms are websites that mediate between buyers (clients) and sellers (freelancers) of remotely deliverable cognitive work (Horton, 2010). One of the biggest freelance platforms of the United States is Upwork. This freelance platform was founded over two decades ago provided a safe environment were freelancers could offer their services before the burst of the current freelance workforce which now makes up 36% of the total US workforce. In order to research the composition of the freelance workforce, Upwork has conducted a report focused on their platform users since 2014. The Upwork Freelance Forward reports conducted by an independent research firm have been providing the freelance industry and scholars from this knowledge domain with accurate market data about this unconventional group of individual workers. However, due to the COVID-19 pandemic, new barriers arose for freelancers. The Upwork Freelance Forward Report (2020) stated a decline of 10% in the freelance workforce for the first time since the conduction of the report. The survey results from the Upwork Freelance Forward reports of 2019 and 2020 were provided by the Upwork research team in order to compare the changes during these two years. This comparison provides a pre-pandemic and midst-pandemic period in order to the 2020's COVID-19 pandemic had impacted the composition of the freelance workforce. By combining literature review study and a self-conducted quantitative study, this thesis tries to predict whether an individual will (dis)continue freelance work during the covid-19 pandemic, using machine learning. The literature review will shine light on the main barriers the economic and societal impact of the pandemic while taking the results of the UFF reports on which to thesis build upon into account. The survey data used in this thesis contains the barriers freelancers would (dis)continue their freelance work in the future. The quantitative research is conducted by comparing the survey data from the two reports and a XG-Boost model is presented to provide an algorithmic method to predict whether a respondent will freelance in the future.

Abstract

This thesis draws on survey data from the annual Upwork Freelance Forward (UFF)[1] reports of 2019 and 2020. The research goal of this thesis is to provide insights in the behavioral drivers behind the decline of the freelance workforce by examining the freelancers' barriers to (dis)continue their freelance work during the Covid-19 pandemic, using machine learning. By comparing the survey data of the two reports, this thesis tries to find the statistical change of the barriers of freelance work. An evaluation of this change during the 2020 pandemic is missing from former research and the UFF report. The timing of the 2020 data collection provides a new insight in the barrier's freelancers experienced during the COVID-19 pandemic. This leads to the research question: *Is it possible to predict whether an individual will (dis)continue freelance work during the covid-19 pandemic, using machine learning?* Indeed, the XG-boost model was able to accurately predict whether a participant decided to stop or continue freelance work. This prediction was based on various barriers among which are: the competitive industry, low and irregular income, difficulty in finding work, and not having an office. This thesis further observed that gender also has an influence on the likelihood of making the decision of stopping pursuing freelance work.

Keywords: *Freelance, workforce, COVID-19 pandemic, Machine Learning, Survey data*

[1] The Upwork Freelance Forward Report of 2020. see: https://www.upwork.com/i/freelance-forward

1. Introduction

1.1. Impact of the covid-19 pandemic on the US Freelance Workforce.

Freelancers are an important, but hidden, part of the total US workforce. Freelance work provides flexible employment arrangements which can be advantageous for both workers, and for organizations looking for on-demand services (Dunn,2017). The US Freelance workforce has grown with almost 50% since 2017 (Stephany et al., 2021). This growth reflects the increased demand for the freelance services required by organizations (Kalleberg, 2003). This increase was reinforced by freelance platforms such as Upwork that connect freelancers and their clients, by providing them with a low-boundary and online marketplace for freelance services (Dunn,2017). In 2020, 36% of the US workforce consisted of freelancers. However, after years of growth, the freelance workforce faced a sudden decline of 10% in the first months of the pandemic, indicating that almost 6 million American freelancers stopped pursuing freelance work during that period. Notably, 88% of the freelancers who stopped during the pandemic wanted to work as a freelancer in the future, which shows that the pandemic significantly impacted the behavior of the freelancers. (Upwork Freelance Forward, 2020). Because the external environment in which freelancers operate is the most important factor determining career success (Van den Born, Van Witteloostuijn, 2013), economic and social disruptions have a substantial impact on their career. This is especially true during the disruptive changes of the 2020's COVID-19 pandemic, adding to the uncertain future for many freelancers (UFF 2020; Dunn, 2020).

1.2. Project context

Upwork is an online freelance platform that connects businesses or organization with a need for freelance competences with the freelancers active on the platform. Freelancers have earned over 2.3 billion US$ in 2020 on Upwork in over 10,000 categories as website & app development, creativity & design, customer support, finance & accounting, consulting, and operations. The platform also provides a cross- social network between freelancers and their clients. Granovetter (1995) demonstrated that trusted social networks can provide individuals with high quality information, business referrals and some social control. The dataset used within this thesis is provided by the Upwork research team and contains survey data from the UFF reports from 2019 and 2020. Quantitative research is conducted to compare the survey data from the 2019 and 2020 UFF reports as the timing data collection provides a scope of both pre- and midst-COVID. Therefore, the focus of this thesis is on freelancer' reasons to (dis)continue freelance work within the context of the COVID-19' economic and social impacts in the United States. Although freelancers are active in many industries (creative, engineering, finance etc.), for the purpose of this thesis, we will focus on high-qualified online freelancers present in the UFF reports.

1.3 The main research questions

The research goal of this thesis is to provide insights in the behavioral drivers behind the decline of the freelance workforce by examining the freelancers' barriers to (dis)continue their freelance work during the Covid-19 pandemic, using machine learning. This thesis adds to the work that recognizes that as the scope and breadth of the global COVID-19 pandemic continues, the implications to freelance workforce and freelance industry are profound (see Related Work).
Based on the introduction, the following main and sub-Research Question (RQ) was formulated:

RQ*:* *Is it possible to predict whether an individual will (dis)continue freelance work during the covid-19 pandemic, using machine learning?*

SRQ 1: *What are the main barriers that predict whether a freelancer will (dis)continue freelance work?*

SRQ 2: *Is there a statistically significant change in the reasons to (dis)continue freelance work 2020, during the covid-19 pandemic, as compared to before the pandemic in 2019?*

1.4 Scientific and societal relevance

Freelancers have been among those hit hardest by the economic challenges of COVID-19 (Atkeson, 2020). It is important to distinguish the two different drivers that are behind the reasons freelancers have to (dis)continue their work. Because the UFF report was conducted during the pandemic, new insights can be created with the argumentation of data collected during the pandemic and shine light on why 10% of the freelancers stopped pursuing freelance work in the period of the conduction of the UFF report of 2020. As new pandemic waves cannot be ruled out (Normile, 2020), there is a need to analyze the human behavioral drivers behind the fluctuations in the freelance workforce, as they are essential to the US economy and society.

The relevance of this topic is evident from the number of studies that were performed in the past 1.5 years. However, they usually rely on statistical techniques such as regression analyses, whereas machine learning algorithms may provide additional insights, given the large datasets in this field. Machine learning approaches may be interesting as they can provide algorithms specifically trained to predict whether a respondent will freelance in the future based on their survey answers. Also, a trained ML model may be able to identify highly relevant variables of the freelance workforce, making it possible to predict whether a freelancer will discontinue their freelance work. These algorithms could be used in related quantitative research.

Finally, the outcome of this thesis can provide valuable business insights for the Upwork platform. Upwork (and parties like it) will be able to get focused data feeds by foregoing long and complicated integrations and modifying business strategies based on the freelancer’s satisfaction patterns. As these satisfaction patterns are influenced by the external environment of freelancer (Van den Born, Van Witteloostuijn, 2013), a ML algorithm that can provide a direct data analysis beneficial for freelance platforms.

1.5 The Findings

The main reasons related to the discontinuing freelance work during the pandemic were increased competition in the industry, low income, difficulty finding work, and not having an office. These results differentiate from the results from the UFF report of 2019. The outcome of this thesis will present new findings about the changed intrinsic motivation of freelancers that contributed to the 10% decline in the freelance workforce.

Chapter 2: Related Work

2.1 The Freelance Workforce

The freelance workforce consists of the skilled professional workers who are neither employers nor employees, and who work on a temporary basis under a contract for a fee and for a range of clients (Schwartz, 2018). In the past decade, freelancing has become an established type of employment in the United States (Atkenson, 2020). The US Freelance workforce has grown with almost 50% since 2017 (Kässi and Lehdonvirta, 2018; Stephany et al., 2021). This growth reflects the increased demand for the freelance services required by organizations. For employers, the hiring of freelancers instead of hiring employees with long term contracts, helps to reduce costs (Manuylova, 2018). Even though securing freelancers is crucial for an organization's growth strategy, organizations often lack the resources and competencies to attract them (Rabino, 2019). The internet has given both employers and employees new opportunities to provide technological and human services without the need for long-term contracts (Wood et al., 2018). The combination of emerged digital innovations and an increased demand for knowledge and resources created a global freelance workforce of 56 million people, 40% of whom live in the US (Stephany et al., 2020). Online freelancing work or services reflects a form of labor that relies on mediated and fully digital interactions between these individual workers and their clients. Many companies in the US use freelancers as a flexible buffer that rapidly can be adjusted in the face of economic changes, protecting employees with a long-term contract (Kalleberg, 2003). It is important to analyze the freelance workforce, as they provide organizations with technological and human solutions by bundling organizational, technological, and human capabilities and competencies (Akhmetshin, 2018).

2.2 Benefits and barriers to freelancing.

Freelancers claim that freelance work provides them with career opportunities, greater work autonomy and a high potential for a good work-life balance, resulting in less boundaries and an individualized, high quality of life (Ashford, 2018). Freelancers often offer their services on online platforms such as Upwork, but also through newspaper ads, by word of mouth or through personal connections (Manuylova, 2018). Akhmetshin (2018) and Kuznetsova, (2018) showed that 46% of their freelance respondents considered that the most important advantage of freelancing work is the opportunity to work remotely and to gain a higher level of employment in isolated communities without the need for geographical relocation. Organizations and businesses in need of freelance services benefit from this advantage as it widens their employee workforce without requiring relocation. Freelance work has increasingly become valuable in many sectors like healthcare, IT, and the creative and cultural industry (Freelance Forward, 2020). Akhmetshin & Kuznetsova (2018) found that a flexible schedule was of significant importance to 45% of the freelancers. Furthermore, the same study found that not having to adapt to a corporate culture and hierarchy is also highly attractive for freelancers. Akhmetshin (2018) and Kuznetsova (2018) also concluded that the collaboration with other freelancers is seen as an important advantage of freelance work. Many freelancers' projects involve cross-industrial collaboration. This collaboration is viewed as a chance for freelancers to expand their professional and social network while developing and exchanging knowledge and competences. Other important advantages are autonomy in the ability to choose one's own projects, career path, working environment and culture without the restraints of corporate demands (Cohen, 2012; Dunn, 2017; Prottas, 2008).

On the flipside, freelancers do not enjoy the same protection as non-freelancers (Fitzpayne, 2019), and freelancers often lack benefits like health care, a pension fund, sick days or leave with pay (Pollack, 2015). Freelance work means increased risks to workers, including an irregular income, not being paid due to lack of work, and the irregular projects demand cause a time overlap between projects, which also causes payment gaps (Ashford et al, 2019; Wood et al., 2019). Also, because of the low barriers to entry, especially the global and digital nature of online freelancing allow for greater competition (Dunn, 2017) and because freelance work is project-based, there is little commitment between the client and the freelancer beyond the specifics of the project's contract (Wood et al., 2019). The PayPal Freelance survey2 (2017) found that the lack of, or a skewed work-life balance is also seen as a disadvantage by freelancers. The experienced work-life balance differs with sector and demographic characteristics, but the main reason for a negatively perception of the work-life balance by freelancers are the working conditions, which require flexibility, long working hours and peak-periods which disturb the work-life balance and consequently make the freelancers' negatively perceive their profession (Suess, 2012). Furthermore, a lack of time to learn new skills is also another downside of freelance work. As freelancers sometimes work on projects unintermittedly, development of knowledge and skills is neglected, in contrast with long-term contract employees who are provided development options and time (Petica-Harris,deGama, & Ravishankar, 2020; Wood, Graham, Lehdonvirta, & Hjorth, 2019). However, working in multidisciplinary and varied projects can also offer great opportunities for development of knowledge and skills. Lastly, having difficult clients also came forward as a disadvantage in many reviewed literatures. The relationships between buyers and sellers can be transactional or embedded; difficult clients are mostly found within the embedded relationship and social disagreements can occur, as difficult clients often create professional and social obstacles for freelancers (Bidwell, 2009). Dealing with strangers in open markets like the Upwork platform is generally riskier than transacting with people one knows, although established social relationships with clients form a better basis for resolving conflicts, difficulties still occur in both long-term and short term-client relations (Cohen, 2012). Freelancers who use virtual communications with their clients rather than face-to-face communications are more likely to be involved into a conflict with clients. The low number or absence of direct encounters between clients and freelancers' results in a higher chance of the client's difficult behavior, disappearance of the client during or after the project, and/or no-payment (Shevchuck, 2012). The disappearance or sudden contract cancellations had a negative impact on the freelancers' experience with their clients. These breaches in contracts not only had an impact on the freelancer's professional life, also their private life had to endure this sudden

2.3 Quantitative Evidence from the Upwork Freelance Forward Report

As this thesis builds on the data and findings of the UFF report, some notable findings that are connected to this research need to be highlighted. Firstly, 10% of the freelancers who paused or stopped pursuing freelance work during the pandemic were typically working in industries where working remotely was not an option. Also, the impact of the pandemic was greater for 'lesser qualified' freelancers, whose jobs were no longer in demand during lockdowns, e.g. cooks, personal care workers or freelance hotel staff and freelancers in the creative industry (Bowles, 2020). Additionally, it was also found that women were more likely to stop with freelance work during the pandemic. Among those freelancers who continued to work, only 61% stated that they were able to work as much or more during the

---

[2] The Paypal US Freelancer insiders report. see: https://www.paypalobjects.com/digitalassets/c/website/marketing/global/shared/global/media-resources/documents/PayPal_US_Freelancers_Insight_Report_Feb_2018.pdf

pandemic year of 2020. On a more positive note, the well-being and mental health of freelance workers was impacted less than that of non-freelance respondents, as they were already more used to a remote working lifestyle, which made them more resilient to lockdown measures.
Although the UFF reports focused on the composition of the freelance workforce, it does not provide insights in the changes of the barriers and motivators that may have caused this decline.

2.4 Freelancing during the pandemic; The economic and societal impact

Initial discussions about the influence of the COVID-19 pandemic on freelancers were characterized by grounded optimism (Atkeson, 2020; McKibben & Fernando, 2020), as the 10% decline after the summer of 2020 has been compensated by new re-entering freelancers and new-entrees. However, it was already expected that a contained global outbreak would have significant global macroeconomic impacts (Atkeson, 2020; McKibben & Fernando, 2020). As the economic impact on the freelance workforce is very versatile, not all economic drivers can be considered. Importantly though, as the COVID-19 pandemic led to a massive rise in unemployment in the United States (OLS, 2020) and in other countries with lower wages, unemployed long-term contractors subsequently (re)entered the online freelance platforms. At the same time, due to lowered confidence in the US economy during the first months of the pandemic, clients were hesitant in hiring more freelancers. This caused increased competition, lower rates, and fluctuations in both demand for freelance services and supply side of the freelance workforce (del Rio-Chanona et al., 2020; Grimov, 2016; Sanchiz, 2021; Schreiber, 2020). This overavailability of online freelancers led to the development of a new social class with a marginal position on the labor-market and a high-risk potential to be victimized by clients making use of the law of supply and demand.

Left unanswered by former literature is how freelancers experienced and reacted to the changes caused by the pandemic. The pandemic exacerbated issues such as irregular income (Dunn, 2020), because a decrease in demand freelance work also causes a lower chance of finding work in sequence (Grimov,2016). Similarly, long-term clients who provided a dependable source of income pre-pandemic stopped projects and did not commission new work. Also, although freelance platforms like Upwork, TaskRabbit and other freelance platforms have drastically reduced the effort and cost of seeking, forming, and terminating work arrangements, searching and retrieving projects is time-consuming task (Benson, 2017). On a more behavioral note, freelancers need to be able to keep up their level of self-discipline without external control (Benson, 2019). This can be significantly impacted as people are now working from home with adapted working hours, many interruptions and distractions from family obligations and rivalry for workspace within the household (Dunn, 2020; Rubin, 2007). Indeed, strained household arrangements caused part of the freelancers to pause or quit their freelance work during the first months of the pandemic (Scheiber, 2020). Former research shows that female freelancers are overall more likely to be negatively impacted by the pandemic in terms of their careers (Guéraut,2021; Blasko,2020; Adams-Prassl,2020).

3: Methods

3.1 The XG-boost model

To answer the main research question, an algorithm was required that could handle the statistics of the survey data, independent variable and dependent variables of choice while

providing a high level of results. Three optional models that could handle multiple independent variables: Linear Regression (MLR), the Ordinal Logistic Regression Model (LRM), and the eXtreme Gradient Boosting (XG-boost) model. MLR is a statistical technique to analyze the relationship between a single dependent variable and several independent variables (Woldemariam, 2017). However, MLR can only be implemented successfully when there are two continuous variables, which was not the case for the selected variables of the UFF dataset, as the answers to questions in the survey were scored on a 5-point Likert scale. The LRM requires linearity between the dependent and the independent variables (Hancock, 2020), an assumption that was not met in the dataset used for this survey.
The XG-boost model was chosen because of the beneficial elements of this model. XG-Boost is a more regularized form of Gradient Boosting. XG-Boost uses advanced regularization (L1 & L2), which improves model generalization capabilities. XG-Boost delivers high performance as compared to Gradient Boosting. Its training is very fast and can be parallelized across clusters. XG-boost is used attempts to accurately predict a target variable by combining the estimates of a set of simpler, weaker models (Hancock,2020). Firstly, XG-boost performs very well on structured and tabular datasets on classification and regression predictive modeling problems. XG-Boost is an open-source implementation of the gradient boosted trees algorithm, which provides code and cost efficiently. Secondly, XG-Boost are trained faster and require less storage space (Hancock,2020). Thirdly, XG-boost algorithm can offer better performance on binary classification problems with a severe class imbalance. Lastly, the model is highly interpretable by looking at the individual trees. As long as the decision tree doesn't have too many layers, it can be interpreted (Woldemariam, 2017). The outcome of XB-boost is the level of importance of the features, also called the importance gain, which implies the relative contribution of the corresponding feature to the model calculated by taking each feature's contribution for each tree in the model. A higher value of this metric when compared to another feature implies it is more important for generating the prediction of the target variable. The gain is the most relevant attribute to interpret the relative importance of each feature (Woldemariam, 2017). The programming language used for the XG-boost model is Python as this is an open-source and easy interpretable programming language with extensive support libraries needed for the training of the XG-Boost model (Ayer & Miguez, 2014).

3.2 Evaluation criteria

Before the evaluation of a model, the possibility of imbalanced classes needs to be taken into account. Fortunately, an advantage of the XG-Boost model is the automatic handling of imbalanced classes. The XG-Boost model provides a large range of hyperparameters which are certain values or weights that determine the learning process of an algorithm (Hancock,2020). The implementation of XG-Boost provides a hyperparameter designed to tune the behavior of the algorithm for imbalanced classification problems; the scale_pos_weight hyperparameter. This hyper parameter has the effect of weighing the balance of positive examples, relative to negative examples when boosting decision trees.

Furthermore, in order to evaluate the performance of a XG-boost model algorithm, a training, validation and testing datasets needs to be created. These sets are created by splitting the original dataset into three parts; train (72%), validation (8%) and the test set (20%), using multiple methods from the sci-kit learn library. The algorithm was then trained the on the training set while the predictions were made on the validation and test set. XG-Boost falls under the category of Boosting techniques in Ensemble Learning. The basic idea behind boosting algorithms is building a weak model, making conclusions about the various feature importance and hyperparameters, and then using those conclusions to build a new, stronger model and capitalize on the misclassification error of the previous model and try to reduce it.

The results of XG-boost model are evaluated on the values of the area under the ROC Curve (AUC), as this was proven theoretically and empirically better than the accuracy metric for evaluating the classifier performance and discriminating an optimal solution during training of XG-Boost models (Bentéjac, 2019). It explains how much the model is capable of distinguishing between classes. An AUC of 0.5 suggests no discrimination; between 0.7 to 0.8 is considered acceptable, and above 0.8 is considered excellent (Hannock, 2020).

4. Experimental setup

4.1. Dataset Description

The datasets used within this thesis are received from Upwork and the UFF reports of 2019 and 2020. The Edelman Intelligence institute collected this data through an online annual survey among N=6001 (n= 2117 freelancers, n=3884 non-freelancer) respondents in 2019, and N=6008 (n=2131 freelancers, n=3869 non-freelancers) in 2020. The respondents were asked a wide range of questions about the outlooks, experiences, satisfaction, and challenges of both freelancers and non-freelance workers. The respondents were weighted by the research team of Edelman Intelligence to ensure that their demographic representation is in line with the US Labor Force Statistics.

4.2. Features

The task of the XG-boost algorithm was predicting whether a respondent will pursue freelance work in the future based on the features found in the dataset. For analysis, a subset of survey questions about the extent of agreement on statements regarding work experience, satisfaction, and challenges rated on a 5-point Likert agreement scale from 0 (low agreement) to 5 (high agreement) were selected by the UFF research team (Table 1). The associated survey questions can be found in Appendix 1. Personal characteristics consisted of gender (0 = Female, 1 = Male), being a freelancer or non-freelancer and age (18-22, 23-38, 39-54 and 55 years old).

Table 1. Features of the XG-Boost model

Features from 5-point Linkert scale questions
Irregular income
Very competitive industry
The income is too low
Difficult clients
It’s hard to be disciplined
Perception of freelancing
Not being paid on time or at all
Hard to find work
Lack of work-life balance
Clients in different time zones
Isolation / working alone

Not having an office
Managing my finances/admin
No time to learn new skills
Features from respondents
Is current Freelancer = 1, Non-Freelancer =0
Gender: Male = 0, Female=1
Age: 18-22, 23-38, 39-54 and 55+
Target Variable
Future freelancer yes= 0, no =1

4.3. Missing values

Because missing values create a reduced statistical power of the model and possibly bias, dealing with them is crucial for the optimal performance of the model. A sparsity-aware algorithm needed to use in XG-Boost to effectively remove missing values from the computation of the loss gain of split candidates. Because XG-Boost is capable of handling the other missing values internally, the algorithm handled the missing answers from the 5-point Linkert scale questions by itself with HistGradientBoosting Regressor which coded missing values as 0. The missing data is handled by minimizing the loss function. The model decided during training whether missing values go into the right or left node. If there are no missing values at training time, it defaults to sending any new missing data to the right node.

4.4. Pre-processing

Firstly, the respondents who agreed with the survey statements (seen in Appendix 1) with a 4 or a 5 on the 5-point Linkert scale, are coded as agreeing with a certain barrier. An example would be that if a respondent would agree with the survey question: '*Irregular income is a reason for me to stop or pause freelancing',* with a 4 or a 5 in the survey data, this respondent would be counted as agreeing with the barrier of 'irregular income'. This generated the agreement of respondents in percentages of every feature. Hereafter, all string data type values were converted to numerical values using label encoders, all Boolean data type values were converted to numerical values in the range [0:1]. String value data includes age groups: 18-22, 22-38, 38-55, 55+, gender: Male, Female. Numerical data types are a requirement for models in order for them to use the data for prediction. To get more summary statistics of the different features in the dataset the describe() method was used on the dataframe. The dataframe generated has no column called futurefreelancer.target because the target column is available in another attribute called futurefreelancers.target. This column was then appended to the pandas dataframe. Furthermore, one-hot encoding was applied on the categorical features before training the model.

4.5. Implementation

After the pre-processing of the data was loaded from the csv. files of the two separate datasets (UFF 2019 & 2020) into Python with the following libraries such as pandas, XG-boost, sci-kit learn and Matplotlib. After loading the data, the input features and the target variable were stored separately. Both datasets were trained to compare the results and find the changes in the feature importance. The model used boosting as a sequential technique which works on the principle of an ensemble. It combines a set of weak learners and delivers improved prediction accuracy. At any instant t, the model outcomes are weighed based on the

outcomes of previous instant t-1. The outcomes predicted correctly are given a lower weight and the ones miss-classified are weighted higher. Note that a weak learner is one which is slightly better than random guessing. The model was build using Trees as base learners ,which are the default base learners, using XG-Boost's scikit-learn compatible API. Firstly, all the relevant variables were added into the XG-boost model. In order to build more robust model, a k-fold cross validation was used where all the entries in the original training dataset are used for both training as well as validation. Also, each entry is used for validation just once. XG-Boost supported this k-fold cross validation via the cv() method. In order to perform the validation, the nfolds parameter was specified to 3, which is the number of cross-validation sets build. This process repeats nfold times until every subsample (fold) has served both as a part of the training set and as a validation set.
Then, the multiple methods from the ski-cit learn library were used to split the data into train (72%), validation (8%) and the test set (20%).

The XG-Boost model also provided a highly efficient implementation of the stochastic gradient boosting algorithm and access to a suite of model hyperparameters designed to provide control over the model training process, seen in Appendix 2. The best performing hyperparameters that were used during the training of the XG-Boost algorithm are provided in Appendix 2. As mentioned in the methods, both datasets consisted of imbalanced classes. By default, this the scale_pos_weight hyperparameter is set to the value of 1.0 and has the effect of weighing the balance of positive examples, relative to negative examples when boosting decision trees. For an imbalanced binary classification dataset, the negative class refers to the majority class (class 0= non-freelancer) and the positive class refers to the minority class (class 1=freelancer). The manual selection of the hyperparameters improved the accuracy of the model as they determent the learning process of the algorithm. As stated earlier, XG-Boost provides large range of hyperparameters. After the training of the model, the feature importance's are collected and the subgroup of the test sets of both the datasets of 2019 and 2020 that are currently freelancer are extracted from the model. This sub-group is used to detect what the most important features are for freelancers to continue freelancer work. Lastly the datasets were converted into an optimized data structure called Dmatrix that XG-Boost supports and gives it acclaimed performance and importance gains. The XG-boost feature importance's were then collected from the model and will be evaluated in the next chapter.

5: Results

5.1 Overall model performance

After the training of this model, the ROC (AUC) performance metric is used to check the performance of the trained model on the test set. In this model, the AUC was used to construct an optimized learning model. Unlike the threshold and probability metrics, the AUC value reflects the overall ranking performance of a classifier. The AUC Curves summarize the trade-off between the true positive rate and false positive rate for a predictive model using different probability thresholds. The ROC curves are appropriate for the evaluation of a model when the observations are balanced between each class. As the XG-Boost used in this thesis provided balanced classes. As the precision-recall curves are appropriate for imbalanced datasets, it is not necessary for the model evaluation.

As the AUC of both training sets are high respectively 0.89 for the 2019 XG-Boost model and 0.91 for the 2020 XG-Boost model, we can conclude that the XG-boost models for both datasets have a high enough performance to draw conclusions on their results.

5.2. Effects of personal characteristics

It was found that there has been a shift in the gender ratio during the pandemic, indicating a relative decline of female freelancers. Furthermore, the ratio of the age groups pre-and post pandemic, are divided differently. Also, it was found that the total freelance workforce of 2020, declined with 10% compared to survey results of the 2019 UFF report, indicating an increase in the non-freelance workforce.

Table 2. Features of the characteristics of the respondents

Features of respondent characteristics	Absolute change per category%	Absolute change per category in%	Absolute change in %	Relative change in %	Feature importance	Feature importance
	2019	2020	2019 vs 2020	2019 vs 2020	2019	2020
Gender (0=female, 1=male)	69% 31%	65% 35%	↓4% ↑4%	↓5,8% ↑12,9%	0.076843	0.079812
Age:(18-22); (23-38); (39-54); (55+)	9,0% 37,0% 30,0% 24,0%	14,13% 33,7% 30% 22,0%	↑5,13% ↓3,3% 0% ↓2%	↑57,0% ↓8,92% 0 ↓8,33%	0.062578	0.068924
Current Freelancer	35,6%	32,0%	↓3,6%	↓10,0%	0.071675	0.074042
Current Non-Freelancer	64,4%	68,0%	↑3,6	↑ 5,6 %	0.081674	0.083682

The XG-Boost model states a changed ratio pre- and post-pandemic in 2020 for the first time in 4 years. Furthermore, the importance gain of gender did increase whilst this value was mostly stable since 2017, based on the statistical ratio of the survey respondents of former years. This also accounts for the age group feature, as the age group ratio changed due to a substantial increase of the younger group (18-22). The importance gain of the age feature also increased post pandemic, which could indicate that age has a higher importance in the XG-Boost model prediction of 2020 compared to the 2019 model.

5.3 XG-Boost feature importance.

Table 2 shows the overall agreement with the barriers in 2019 and 2020, as well as the importance gain of the features. The importance gain of the XG-boost model implies the relative contribution of the corresponding feature to the model calculated by taking each feature's contribution for each tree in the model (Bastidas, 2021).

Table 2. Main barriers for the respondents to stop working as a freelancer in order of importance, and the feature importance resulting from the XG-boost model.

Features of the barriers	Respondent agreement in % 2019	Respondent agreement in % 2020	Absolute change in % 2019 vs. 2020	Relative change in % 2019 vs. 2020	Feature importance 2019	Feature importance 2020
Irregular income	39%	46,20%	7,2%	18,4%	0.062578	0.063265
Very competitive industry	22,10%	29,10%	7,00%	↑ 31,67%	0.080342	0.089153
The income is too low	21,10%	24,30%	3,20%	↑ 15,56%	0.080342	0.088942
Difficult clients	18,10%	18,30%	0,20%	↑ 0,01%	0.00000	0.00000
It’s hard to be disciplined	19,30%	15,20%	-4,10%	↓ -21,24%	0.00000	0.00000
Isolation / working alone	19,20%	10,40%	-8,80%	↓ -45,83%	0.066842	0.046136
Perception of freelancing	16,10%	15,10%	-1,00%	↓ -6,21%	0.036942	0.038938
Lack of work-life balance	15,10%	15,20%	0,10%	↑ 0,67%	0.00000	0.00000
Clients in different time zones	14,10%	13,80%	-0,30%	↓ 2,13%	0.00000	0.00000

Not having an office	12,10%	23,70%	11,60%	↑ 95,86%	0.056841	0.0868957
Not being paid on time or at all	9,30%	9,80%	0,50%	↑ 5,38%	0.00000	0.00000
Hard to find work	8,50%	14,20%	5,70%	↑ 67,01%	0.077836	0.087977
Managing my finances/admin	8,10%	7,50%	-0,60%	↓ 7,41%	0.00000	0.00000
No time to learn new skills	5,10%	4,20%	-0.90%	↓ -17,65%	0.00000	0.00000

The results show that what people pre-pandemic indicated to be (the top 5 of) important factors influencing their decision to freelance are, in order, having an irregular income, the competitiveness of the industry, a low income, difficult clients, and difficulty with discipline. In 2020, the top self-reported barriers were also the irregular and low income, as well as the competitiveness of the industry, albeit that the number of participants who agreed with these statements increased significantly as compared to 2019. However, in 2020, the number of participants who indicated that the lack of an office was a barrier for freelancing skyrocketed both in mentions and in importance gain.

However, these numbers indicate what participants themselves indicate to be important barriers for their decision. However, looking at the feature importance results, which barriers predict their actual behavior in the moment can be different. The barriers of low and irregular income and the competition are not only often agreed with, but they also significantly predicted the participants' decision to freelance or not. In contrast, the barriers regarding difficulty with staying disciplined and having to deal with difficult clients provided no predictive power, even though freelancers saw them as important disadvantages of freelance work. In other words, what participants deem to be important, may not actually be a predictor of their actual behavior. Difficult clients for example were deemed an important barrier for freelancers, however, it did not have a high importance gain, indicating a low importance in the predicting the target variable. Similar results were found for a lack of work-life balance, and not begin paid on time. Also, both pre, - and midst pandemic, not many participants regarded the administrative tasks and a lack of time to learn new skills as very important barrier for their freelance status. In contrast, the number of participants who agreed that the difficulty in finding work,

Chapter 6: Discussion

The research goal of this thesis was to assess the main barriers to freelance work and the effect of the covid-19 pandemic on these barriers. Overall, the aim of this study was however to assess whether it was possible to predict whether an individual would (dis)continue freelance work during the covid-19 pandemic, using machine learning. Indeed, the XG-boost model was able to accurately predict whether a participant decided to stop or continue freelance work.

6.1. The effect of personal characteristics.

Firstly, personal characteristics influence the likelihood of freelancing. Participants who were freelancers at the time of the study were more likely to freelance in the future, possibly because non-freelancer face increased barriers to enter the freelance workforce. More interestingly, it was found in the results of the XG-Boost model that there was a decline of female freelancers, while this group has been stable since 2017 (OLS,2017;2020). This corresponds with the results of the literature review (Dunn, 2020) that reported that female freelancers are more likely to stop pursuing freelance work during the pandemic due to caregiving tasks (Dunn,2020). Although, the results from this thesis do not show a substantial decline as to be expected from the results of the literature, the sudden change in gender ratio, in favor of the male group, could indicate that there were more women among the stopped freelance group. Studies from other research fields also show that women are overall more likely to be negatively impacted by the pandemic in terms of their careers (Guéraut,2021; Blasko,2020; Adams-Prassl,2020). However, drawing strong conclusions on the decline of female freelancers is difficult with the results of this thesis as this decline also be caused by a higher ratio of male freelance entrees (UFF,2020).

6.2. Pandemic independent predictors of the decision to freelance.

More interesting are the barriers, which are a potential target for change interventions. The data comparison from the two reports in 2019 and 2020 with the help of the XG-Boost model showed the pandemic’s influence impacted the barriers freelancers had to (dis)continue freelance work. Three barriers or drawbacks associated with freelancing, namely low income, irregular income, and high competitiveness, were not only mentioned by many participants as important factors for their decision to freelance, but they also allowed the author to significantly predict the decision to freelance both before and during the pandemic. However, the pandemic did increase the importance of these barriers. This is in line with the previous literature (PayPal, 2017; Dunn,2020; Ashford et al, 2019; Wood et al., 2019; Akhmetshin & Kuznetsova, 2018). The findings that a lack of discipline, clients in different time zones, problems with getting paid on time, and a lack of time to learn new skills weren't important features are contradictory to expectations based on previous literature (Paypal,2017; Grimov,2016 ;Cohen, 2012). Two other drawbacks of freelancing, having to deal with difficult clients and a lack of work-life balance, were to be found similar before and during the pandemic. The pandemic also didn't really change the participants’ perception of freelancing, but this feature also has a relationship with the decision to continue freelance work although less substantial than the other features.

6.3. Pandemic dependent predictors of the decision to freelance.

Also in line with the previous literature are that the pandemic increased the importance of the barrier ‘difficulty in finding work’ (Paypal, 2017; Akhmetshin & Kuznetsova, 2018) and the lack of office space. In 2020, it was already reported that especially freelancers who were not working remotely before the pandemic were twice as likely to pause freelancing compared to those working remotely (Freelance Forward, 2020). Interestingly, it was also found that the ‘lonely’ or ‘isolated’ aspect of freelancing suddenly became less important during the pandemic dan before the pandemic, probably as the lockdowns (Sanchiz, 2021). This was shown both as a lower overall agreement with this barrier, as well as a lower feature importance score.

6.4 Limitations of this research

Although this paper provides a valuable framework for the influence the pandemic had on the freelance workforce It is important to note the limitations of the research involved in this thesis. The first limitation is the scope of the paper as it focuses on the Freelance workforce of the United States and therefore results may not be applicable to freelance workforces of other countries. Also, a selection of features and matching survey questions were considered when evaluating the reasons for freelancers to stop or pause working. Furthermore, some freelance characteristics that also could influence the decision of continuing freelance work during and after the pandemic like; industry they were working in, type of work, being a full-time or part-time freelancer, years of freelance activity, and educational level were not taken into account and the model as the scope of this paper is limited. Another limitation of this thesis is the proposed machine learning model. This model could be limited in applicability on the machine learning problems as the model is trained on this specific dataset and selected features. Also, the model is trained on features chosen by the author and could only provide insights for the chosen features and not for features that were not taken into account in this paper. Another drawback of the model is the fact that the model uses the current freelancer status as another predictor is a drawback. It would have been preferable if the model would have performed equally as good for freelancers, as for non-freelancers. This way it would have been possible to report if the two groups had very different barriers. This could show how to retain current freelancers, and how to help non-freelancers to become a freelancer. Lastly, there are limitations on drawing conclusions on the changed age and gender results due to the contradictory lower result of the decline of female freelance workforce and possibility of misleading results.

6.5 Contribution to the existing framework

Building from the 10% decline of the US Freelance workforce found in the UFF 2020 report, there was a need for a better understanding of the barriers freelancers had to (dis)continue Freelance work. This thesis contributes to the knowledge domain of the freelance industry as well to the future UFF reports and COVID-emerging literature. Taken together, findings suggest additional arguments of the theoretical framework for the relationship between the COVID-19 pandemic and the 10% decline of the freelance workforce. The results of the XG-Boost algorithm established the importance's of the barriers freelancers experienced, providing the current theoretical framework with insights in which barriers are the most influential on the decision to (dis)continue freelance work. Furthermore, this thesis provides an XG-boost machine learning algorithm that predicts the continuity of freelance work based on these barriers that could be implemented in the UFF research of 2022 or other similar data research.

Chapter 7: Conclusions

7.1 The main findings

The COVID-19 pandemic’s influence continues to unfold around the world and challenges most societies’ ability to deal with socio-economic barriers that arose during this disruptive period. This thesis aimed to assess if it is possible to predict whether an individual will (dis)continue freelance work during the covid-19 pandemic, using machine learning. With the help of machine learning, the changes of socio-economical barriers freelancers experience during the pandemic could be detected. It was shown that there is indeed a statistically significant change in the reasons to (dis)continue freelance work 2020, during the covid-19 pandemic, as compared to before the pandemic in 2019. It was also found that personal characteristics influence the likelihood of freelancing in the future. The decline of female freelancers, and the higher likelihood of the continuing with freelance work as a freelance participant indicated the influence of these personal characteristics. The main barriers of freelance work found are the perceptions of low and irregular income, as well as the competitiveness of the industry. Although these barriers are also found as drawbacks in pre-pandemic literature, indicating an independence from the pandemic, these barriers did increase in importance during this disruptive period. Furthermore, during the pandemic, a perceived sense of isolation that can be associated with online freelance work suddenly became way less important, whereas the lack of an office space and the difficulty of finding work became a lot more important. These findings are not difficult to place during the pandemic, when isolated from face-to-face contact with coworkers and novel household arrangements.

7.2 The future of the freelancer workforce

The freelance work' experiences during the pandemic highlight the risks and opportunities of flexible employment relationships and foster a deeper understanding of the potential role of different stakeholders such as the Upwork platforms and the clients that hire freelancers. The insights of this thesis that are offered to scholars of work can accommodate the structure of precarious work on better understanding of the barriers the US Freelance workforce experienced during the pandemic. The changed economic and social barriers of the pandemic mentioned in this thesis provide scholars seeking to draw their expertise on the freelance workforce with a new academic scope. The results could also be used in the UFF report of 2022 by continuing with in-depth research on the socio-economic barriers freelancers experience. Furthermore, the UFF research team could use the machine learning algorithm as a starting point for further machine learning research on future surveys. As the new future of work post-COVID will expand the possibilities for freelancing, the socio-economic barriers of this workforce will continue to play a crucial role in having a successful freelance career. There is a need for policymakers considering these issues, as they think about how to promote freelance' satisfaction during the pandemic. The focus on diminishing these barriers for freelancer should become increasingly important for both their clients as for society who reap the rewards of this valiant workforce. Ideally, this leads to the development and enactment of new policies and regulations that support freelance workers overcome the socio-economic barriers that come with their occupation. As freelance work becomes more precarious during pandemic periods, how do we translate or interpret the current and extensive knowledge regarding the societal and economic barriers driving the US freelance workforce?

Acknowledgements

Throughout the writing of this thesis I have received a great deal of support. First, I would like to thank my supervisor, Dr E.M.J Huis in ‘t Veld, whose expertise was invaluable in formulating the research questions and methodology. Your insightful feedback pushed me to sharpen my thinking and brought my work to a higher level. I would also like to thank Judita Rudokaitė who gave me great guidance and insights in the technical part of my thesis.

I would also like to thank the Upwork research team for providing me with the Freelance Forward Dataset. You provided me with the tools that I needed to choose the right direction and successfully complete my thesis.

References

Adams-Prassl, A., Boneva, T., Golin, M., and Rauh, C. (2020). Inequality in the Impact of the Coronavirus Shock: Evidence from Real Time Surveys. *J. Public Econ.* 189, 104245. doi:10.1016/j.jpubeco.2020.104245

Akhmetshin, E, et al. "Freelancing as a type of entrepreneurship: Advantages, disadvantages and development prospects." *Journal of Entrepreneurship Education* 21.2 (2018): 1528-2651.

AON. COVID-19 Outbreak: Supporting Employee Wellbeing (2020). Online available at https://www.aon.com/getmedia/12ef3be1-0d77-477a-86bd-0ef96372cbd0/Supporting-Emp loyee-Wellbeing-During-The-COVID-19-Outbreak-2020-03-27.aspx. Retrieved on 21/12/2021

Ayer, V & Miguez, S. (2014). Why scientists should learn to program in Python. Powder Diffraction. 29. S48-D64. 10.1017/S0885715614000931].

Atkeson, A. (2020). What will be the economic impact of COVID-19 in the US? Rough estimates of disease scenarios (Working Paper 26867). National Bureau of Economic Research.

Bastidas, O & Zapirain, A & Begoña, Z & Totoricaguena, S. (2021). Feature analysis and prediction of complications in ostomy patients based on laboratory analytical data using a machine learning approach. 1-8. 10.1109/BIOMDLORE49470.2021.9594427.

Benson, A & Sojourner, A & Umyarov, A. (2019). Can Reputation Discipline the Gig Economy? Experimental Evidence from an Online Labor Market. Management Science. 66. 10.1287/mnsc.2019.3303.

Bentéjac, C & Csörgő, A & Martínez-Muñoz, Gonzalo. (2019). A Comparative Analysis of XGBoost.

Blasko, Z., Papadimitriou, E., and Manca, A. R. (2020). *How Will the COVID-19 Crisis Affect Existing Gender Divides in Europe, EUR 30181 EN,*. Luxembourg: Publications Office of the European Union. doi:10.2760/37511

Blau, F., & Kahn, L. (2016). The Gender Wage Gap: Extent, Trends, and Explanations. Technical Report 21923. National Bureau of Economic Research. 1–77 pages. https://www.nber.org/papers/w21913.pdf

Caballero, G.A. (2017). Responsibility or autonomy: Children and the probability of self-employment in the USA. Small Business Economics, 49(2), 493-512.

Cohen, N. (2012). Cultural Work as a Site of Struggle: Freelancers and Exploitation. Triple C: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society. 10. 10.31269/vol10iss2pp141-155.

Duggan, J., Sherman, U., Carbery, R., McDonnell, A.: Algorithmic management and app-work in the gig economy: a research agenda for employment relations and HRM. Human Resource Manage. J. 30(1), 114–132 (2020)

Dunn, M. (2017). Digital Work: New Opportunities or Lost Wages? American Journal of Management, 17(4), 10–27.

Dunn, M., Stephany, F., Sawyer, S., Munoz, I., Raheja, R., Vaccaro, G., & Lehdonvirta, V. (2020). When motivation becomes desperation: Online freelancing during the Covid-19 pandemic.

Van den Born, A., Van Witteloostuijn, A. (2013). Drivers of freelance career success. Journal of Organizational Behaviour, 34: 24-46.

Foone E., Vincent, N. & Hecht B., & Gerber E. (2018). Women (Still) Ask For Less: Gender Differences in Hourly Rate in an Online Labor Marketplace. Proc. ACM Hum.-Comput. Interact. 2, CSCW, Article 53. https://doi.org/10.1145/3274322

Guéraut, E & Lambert, A & Girard, V. (2021). Socio-economic Impacts of COVID-19 on Working Mothers in France. Frontiers in Sociology. 6. 10.3389/fsoc.2021.732580.

Grimov, O. (2016). Main features of freelancing as a non-standard form of employment. Economic Annals-XXI. 157. 79-81. 10.21003/ea.V157-0024.

Hancock, J & Khoshgoftaar, T. (2020). CatBoost for big data: an interdisciplinary review. Journal of Big Data. 7. 10.1186/s40537-020-00369-8.

Horton, J. J. (2010). Online Labour Markets. In A. Saberi (Ed.), Internet and Network Economics (pp. 515–522). Springer: Berlin Heidelberg. https://doi.org/10.1007/978-3-642-17572-5_45.

International Labour Organisation. COVID-19 and theWorld ofWork (2020). Online available at https://www.ilo.org/wcmsp5/groups/public/@dgreports/@dcomm/documents/briefingnote/wcms_745963.pdf. Retrieved on 21/6/2020

Kalleberg, A. L. (2003). Flexible firms and labour market segmentation: Effects of workplace restructuring on jobs and workers. *Work and occupations*, *30*(2), 154-175.

Kniffin, K (2020). *COVID-19 and the Workplace: Implications, Issues, and Insights for Future Research and Action*. 10.31234/osf.io/gkwme. Dyson School of Applied Economics and Management, SC Johnson College of Business.

Kuznetsova, L.M. (2018). Problems of self-employment of the population in the Russian Federation. *Science and education without borders: basic and applied researches, 7,* 131-136.

Lehdonvirta, V. (2018). Flexibility in the gig economy: managing time on three online piecework platforms. *New Technology, Work and Employment*, *33*(1), 13-29.

Manuylova, A. (2018). *Freelance goes a digital way*. Labor market monitoring. Kommersant newspaper. Retrieved from http://www.kommersant.ru/doc/3553834

PayPal. (2017). U.S. Freelancer Insights Report. Retrieved from https://www.paypalobjects.com

Pollack, J & Forster, W (2015). Promotion- and Prevention-Focused Networking and Its Consequences for Entrepreneurial Success. Social Psychology and Personality Science. 6. 3-12. 10.1177/1948550614543030.

Prottas, D. (2008). *Do the self-employed value autonomy more than employees?* Research across four samples. Career Development International. 13. 33-45. 10.1108/13620430810849524.

PWC. Returning to the Workplace after COVID-19 (2020). Online available at https://www.pwc.com/us/en/services/governance-insights-center/library/covid-19-returning-workplaceboards. html. Retrieved on 21/12/2021

Rubin, B (2007). Time–Work Discipline in the 21st Century. Research in the Sociology of Work. 17. 1-26. 10.1016/S0277-2833(07)17001-1.

Sanchiz, C, Carvajal, M & González-Esteban, J (2021) Survival Strategies in Freelance Journalism: An Empowering Toolkit to Improve Professionals' Working Conditions, Journalism Practice, DOI: 10.1080/17512786.2021.1929414

Schwartz, D. (2018). Embedded in the crowd: Creative freelancers, crowdsourced work, and occupational community. *Work and Occupations, 45*(3), 247-282.

Shevchuk, A & Strebkov, D. (2012). Freelance Contracting in the Digital Age: Informality, Virtuality and Social Ties. SSRN Electronic Journal. 10.2139/ssrn.2172702.

Spurk, D., Straub, C.: Flexible Employment Relationships and Careers in Times of the COVID- 19 Pandemic (2020)

Suess, S & Sayah, S. (2011). Work-Life Balance of Freelancers between Reality and Ideal Conception - An Explorative Approach. Zeitschrift für Personalforschung. 25. 247-268. 10.1688/1862-0000_ZfP_2011_03_Suess.

Wood, A.J., Lehdonvirta, V., & Graham, M. (2018). Workers of the internet unite? Online freelancer organisation among remote gig economy workers in six Asian and African countries. *New Technology, Work and Employment*, *33*(2), 95-112.

Woldemariam, K. (2017). Multiple Linear Regression using Python Machine Learning. 10.13140/RG.2.2.35706.59849.

Appendices

Appendix 1

Survey Questions with a 5-point Linkert scale

Survey Questions used as feature
Irregular income is a reason for me to stop or pause freelancing
The competitive industry is a reason for me to stop or pause freelancing
The low income is a reason for me to stop or pause freelancing.
The difficult clients are a reason for me to stop or pause freelancing
Lack of self-discipline is a reason for me to stop or pause freelancing.
My changed perception of freelancing is a reason for me to stop or pause freelancing.
Not being paid on time is a reason for me to stop or pause freelancing.
Difficulty of finding work is a reason for me to stop or pause freelancing
The lack of work-life balance is a reason for me to stop or pause freelancing.
Having clients in different time zones is a reason for me to stop or pause freelancing
The isolation / working alone is a reason for me to stop or pause freelancing.
Not having an office is a reason for me to stop or pause freelancing.
Managing my finances/admin is a reason for me to stop or pause freelancing.
Having no time to learn new skills is a reason for me to stop or pause freelancing.

Appendix 2: The XG-boost hyperparameters

Parameter	Role/Function	Range	Default
N_estimators	The number of trees boosted	1 - inf	2000
Learning rate or schrinkage	Parameter that controls the weighting of new trees added to the model so helps modify the update rule.	0 - 1	0.03
Colsample.by tree	The fraction of features to be evaluated at each split	>0 - 1	1
Scale POS weight	dealing with imbalanced classes)	positive class/negative class = future freelancers count / future non-freelancer count	1
Seed	Parameter that determine the path of trees that focus on different part (e.g. subset of columns) of training data.	0 - inf	1
Max depth	The length of the longest path from the tree root to a leaf.	0 - inf	3
N_jobs	The number of parallel threads used to run XG-boost.	0 - inf	-1
Gamma	The minimum loss reduction. The higher this value, the shallower the trees.	0 - inf	5
N_folds	The number of cross-validation sets build	0-inf	3