Will OECD Countries Achieve Significant Drop in NEET Levels by 2020? Forecasting NEET in OECD Countries

Kerem Tugberk Capraz
5028475

This research report is submitted as term paper for the class
Sustainable Development Goals - 2020, Summer Semester, Lectured by
Prof.Dr.Jürgen Schupp

Sociology: European Societies
Freie Universität Berlin
Germany
30.07.2020

Will OECD Countries Achieve Significant Drop in NEET Levels by 2020? Forecasting NEET in OECD Countries

Kerem Tugberk Capraz

July 30, 2020

Abstract

This paper investigates the potential levels of NEET (youth not in employment, education or training) in 2019 and 2020. The forecasting tasks have been carried out by Catboost gradient boosting algorithm. Historical records between 1997 and 2018 of Four predictors - Unemployment rate, inflation rate, trade in goods and services, and, gdp growth rate - were used for training the model. Following the model training, OECD and IMF projections of the mentioned indicators for the year 2020 were plugged into the trained model to make predictions for NEET values in 2020[1]. The projections made by the model show a significant increase in the mean NEET values in the OECD countries in 2020. Compared to 2015 - reference year for Sustainable Development Goals - the vast majority of the OECD countries are forecasted to fail on achieving the target of having a significant drop in NEET values by 2020.

Keywords **:** $NEET, Forecasting, Catboost, SustainableDevelopment$

1 Introduction

The OECD indicator, youth not in education, employment or training (NEET hereinafter) is a negatively defined concept. Rather than defining characteristics of the target group, it underlines what are not the characteristics of that specific group. The term youth in the indicator statement refers to adults aged between 15 and 29. In OECD's definition, the term education covers the activities of "attending part-time or full-time education, but excludes those in non-formal education and in educational activities of very short duration."(O.E.C.D., 2020c). Within the scope of NEET indicator, the term employment refers to being "in paid work for at least one hour in the reference week of the survey or were temporarily absent from such work." (O.E.C.D., 2020c).

[1]The raw data analysis made this paper available can be found at: `https://github.com/tugberkcapraz/Predicting_NEET_2020/blob/master/NEET_new.ipynb` .

Along with retirement option for senior citizens, the above mentioned activity areas constitute the majority of the life course of human beings. For OECD, being a NEET bears a risk of social exclusion. Also, from a normative standpoint, these 3 activity areas also bring a line of personal development path and a way to prevent social exclusion. Former British Prime Minister Tony Blair phrases this normative standpoint as follows:

> "The best defence against social exclusion is having a job, and the best way to get a job is to have a good education, with the right training and experience."(Unit, 1999, p. 6)

Both Blair and OECD define NEET people as socially excluded. This vague wording can be brought into more practical level with the following question: What are those young people doing if they are not involved in either one of educational, occupational or training related activities? Since there is almost no data, we don't have the exact answer to this question. However, it's not hard to speculate that NEET people, in the most optimistic case, reside in the realm of informal economy. From a more pessimistic perspective, NEET can be associated with criminal activities, political and/or religious extremism as it's often portrayed in media organs. Even this alone, makes NEET an important social matter . However there are more reasons to tackle NEET and keep it low if not eradicate it all. According to Steve Coles, a longitudinal research carried out by Scottish government revealed that being NEET is associated with a higher risk of both poor physical and mental health after 10 and 20 years. Same report also highlights that NEETs are more likely to take jobs that require low-skills. Similarly, the report revealed that NEETs' educational attainment also remains low in comparison with their non-NEET peers through the time (S. Coles, 2017). Ryan calls this sort of harm caused by prolonged unemployment status to people's chances in the job market in later stages of their lives as "scarring effect" (Ryan, 2001). This effect makes handling the NEET as financially very costly in the long term as it increases welfare dependency of those people (B. Coles, Godfrey, Keung, Parrott, & Bradshaw, 2010).

As Holte points out, it's also important to note that being NEET is experienced as a dynamic process rather than being a frozen state by the young people (Holte, 2018). There may be several ins and outs during a NEET's lifetime. As Denny and colleagues note this dynamic process brings about the entrepreneurial prospect which is perhaps the only positive aspect of NEET (Denny, Hazenberg, & Seddon, 2010). That being said, it should also be noted that this entrepreneurial prospect might not be equally shared by all NEETs for various structural reasons. However, the automatisation of the low skilled jobs through the widened usage of artificial intelligence can threaten the labor market entry chances of NEETs altogether.

2 Theoretical Frame

In order to better predict the future levels of NEET, the first thing to do is to select a set of predictors that have direct or indirect causal links to the NEET. A brief review of the literature on NEET reveals the multilayered and heterogeneous nature of the phenomena at hand. I will divide the literature into two main branches. The first one is the micro-level causes of the NEET. This branch of literature on NEET consists of studies which suggests or enquires the existence of a causal line between micro level indicators and NEET status. For example, ethnicity (MacMillan, Gregg, & Britton, 2012), Migration background(Krause & Liebig, 2011; Liebig, Kohls, & Krause, 2012), gender (Strand, 2007), social status and demographic structure of the family that the NEET person raised in (Williamson, 1997) , mental and physical Health (Rodwell et al., 2018) can be viewed among the first branch of literature. These causes pinned by the researchers point out to individual level or group level characteristics that are believed to affect NEET status. Second branch of the literature is the macro level causes of NEET. The studies within this branch approach the issue from an economic perspective. It should be noted that this branch of literature is limited. The macro level studies mostly deal with either one of the terms embodied in OECD's NEET definition.

There is a wide array of Macro level explanations for unemployment, youth unemployment, education and vocational training. The studies concerned with labor market transition can also be viewed as complementary literature for approaching the NEET from a macro level perspective. The limited macro level studies directly focusing on NEET mostly deals with the impact of 2008 crisis on the NEET levels (Bruno, Marelli, & Signorelli, 2014; Dietrich, 2013; Caroleo, Rocca, Mazzocchi, & Quintano, 2020; Kelly & McGuinness, 2013; Rodriguez-Modroño, 2019). A closer glance through this limited literature displays array of potential macro level causes for NEET status. These are business cycle, incidence of temporary employment, educational attainment rates among young people. Among these, only the indicators that can be clustered under business cycle variable have 2020 projections. The latest data points available for the remaining variables are from 2019. This limits the potential predictors that can be added to this study. For this limitation, current study selects 2 main predictors. They are the "business cycle" and the "monetary vs labour force" variables.

The studies found out that NEET rates are really sensitive to business cycle fluctuations. The rationale for this oversensitivity is the last in first out (LIFO) approach adopted by the companies. In the event of a sharp and rapid decline in business productivity LIFO is triggered (Bradley, Migali, & Navarro Paniagua, 2019). As the NEET people, not surprisingly, in most cases are the latest entrants of economic production, during a rapid down-turn they become the first ones to lose their jobs. To measure business cycle, this study employs two OECD indicators which have 2020 projections. The first one is the gdp growth rate and second one is trade in goods and services in the export. Even under the single-hit coronavirus projections of OECD both two indicators are expected to fall sharply (O.E.C.D., 2020a, 2020b). The fall in gdp can reflect the narrowing down of a national economy as a whole. The decline in trade in goods and services in the export is a complementary indicator to gdp. It measures the percentage of the exports within the gdp. With this variable added, the model can differentiate the export oriented economies and the rest.

For the monetary policy vs labour force variable, the literature review brought no direct relationship of it to the NEET. To be more exact with the wording, what is meant by monetary vs labour force variable is the argued trade-off between unemployment and inflation. According to original and more conservative interpretation of the Phillips Curve, there is a negative relationship between unemployment and wage inflation which in turn triggers price inflation (Phillips, 1958). Later on this argued stable trade-off has been theoretically widened and revised by Milton Friedman. According to Friedman's short-run Phillips curve depiction, there is no direct and stable association between unemployment and inflation at all ranges; rather there is a relationship between "natural rate of unemployment" and inflation (Friedman, 1977). In either two interpretations of it, these two indicators do not directly relate to the NEET, however, I shall add those two indicators to the model in order to make model to learn characteristics of each economy with respect to those indicators during both the normal times and the crisis times.

3 Methods and Data Operations

As noted above, there are two variables, business cycle and monetary policy vs labour force, selected for this study. The operationalization of the variables in the current study is as follows:

Variable	Indicators
Monetary policy vs Labour Force	-Unemployment rate -Inflation rate, average consumer prices
Business Cycle	-Gdp growth per annum -Trade in goods and services in Export

Table 1: Table 1: Variables and Operationalization

Both indicators for monetary policy vs labour force variable are derived from the IMF (I.M.F, 2020a, 2020b). The source of indicators for the Business cycle variable is OECD (O.E.C.D., 2020a, 2020b). As it's explained above, the most important point in variable selection is that the availability of the data for 2020. The indicators like educational attainment or temporary work are not included in this study because of this limitation. Since the year 2020 is associated with coronavirus pandemie, thus expected to break up with the ongoing trends, any form of missing data imputation is not healthy for the analysis.

The analysis has been carried out for 33 OECD countries. The OECD countries that fulfil the criteria of having data points for at least 10 years for all indicators and the countries which have no missing values between the years of 2008 and 2013 are kept in the scope of this study.

For the forecasting NEET values in 2020, this paper uses Catboost gradient boosting algorithm. Gradient boosting algorithms are, in general, far better at prediction tasks than their traditional statistical counterparts like linear and logistic regression. What makes them superior is that they are not restricted with linearity assumptions. They can detect non-linear relationships without explicitly commanding them. However, this performance boost comes with a price. The gradient boosting algorithms are called blackbox models (Breiman et al., 2001) . The reason that they are called blackbox is that these tools do not provide parameters which match the input to the output. Given that the task is to make predictions, rather than exploring the relationships between the variables, the trade-off between explainability and performance seems fair here.

Catboost algorithm, iteratively, reshuffles the training set and builds decision trees on those pairs of distinct training sets. The algorithm calculates a loss function, which is basically the difference between the predicted value and observed value, to find the right direction towards better predictions. To prevent overfitting, the algorithm fits each model to a pair of the dataset (test set) which she has never seen before. By the end of this iterative process, the model which yields the optimal results in both datasets becomes the final model (Prokhorenkova, Gusev, Vorobev, Dorogush, & Gulin, 2018).

In order for Catboost to work, the dataset is split into 2 parts by the researcher making sure that each set contains data points for every year in the whole dataset. By this way a potential year wise non-representative splitting is prevented. The 80 % of the dataset spared for training the model and the remaining 20 % left out for testing purposes. The figure 1 and figure 2 respectively show model predictions against the observed values in both training and test data sets.

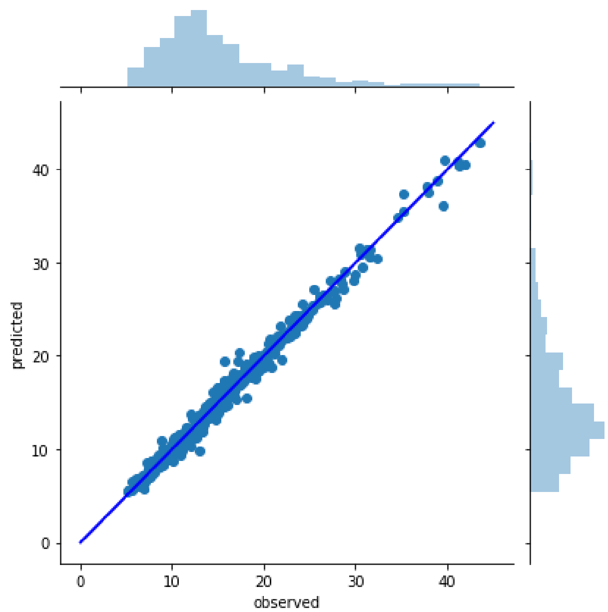

Figure 1: Observed values vs predicted values during the training process.

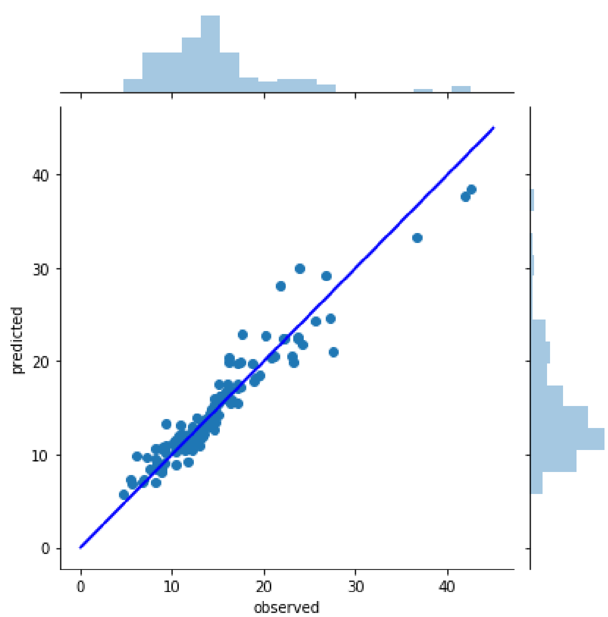

Figure 2: Observed values vs predicted values during the test process.

As it can be seen above, the model is slightly overfitting. This overfitting can be quantified by using different goodness-of-fit metrics and coefficient of determination. The following table represents model criticism metrics:

Metrics	Data set	
	Training	Test
Mean Squared Error	0.5284799520979059	3.2324659384827372
Median Absolute Error	0.39411046074582323	0.8716680832404675
Mean Squared Log Error	0.0022940548754641784	0.011976161317056177
R^2	0.988196578894564	0.916430619829086

Table 2: Model Criticism

Since the machine learning framework is more adopted and led by industry practitioners, there is no widely accepted standard of acceptable values. However, the low differences between mean squared error, median absolute error, and mean squared log error values of model in both training and test sets suggests that model's out of sample performance can be considered acceptable. The common metric of R^2, coefficient of determination, is concerned with the explanatory capacity of the indicators with respect to the variance of the dependent variable. Roughly 7 % drop in R^2 metric can be considered a big drop in the explanatory capacity if the rate itself is low. Nevertheless, as the out of sample explanatory capacity remains above 90 % this drop can be ignored.

Lastly, although blackbox models are not interpretable by themselves, there are workarounds emerging such as Lime (Ribeiro, Singh, & Guestrin, 2016), Saabas (Saabas, n.d.), and Shap (Lundberg et al., 2020) methods. Shap(Shapley Additive Explanations) is an explainable artificial intelligence method borrowed its core from Shapley values from cooperative game theory. It breaks down every single prediction made by the model into the effects of each predictor by following the decision tree of the model and replacing the values with counterfactuals in different perturbations (Lundberg et al., 2020). The following graph shows how each and in which direction the predictors contributed to the models predictions.

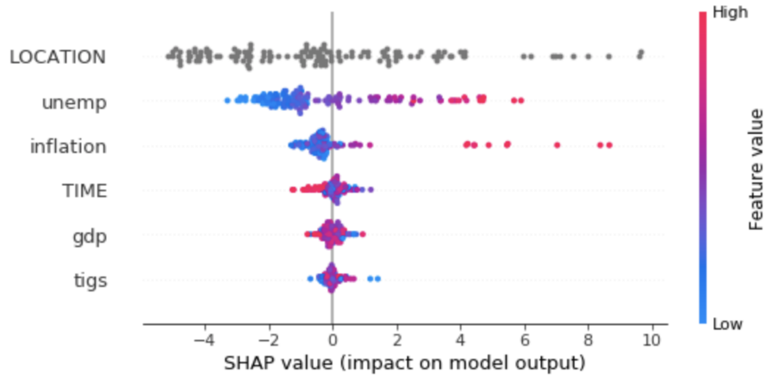

Figure 3: Summary Plot of Shapley Values.

The shapley values are presented in log-odds scale. The model calculates a universal expected value for each year's NEET value. The shapley values represent how much a predictor contributed to the difference between model expectancy and the model prediction.

The location is greyed out because Catboost treated it as a categorical variable. It became the most important variable for the model to make the predictions. Following that, the model relied more on unemployment and inflation indicators. The higher observed values of each indicator (marked red) have pushed the predictions above the expected value. As for business cycle variable, both indicators remained of relatively low importance for the model. Model, couldn't get a clear signal solely from those indicators. The Time indicator is treated as a standard numeric variable by the Catboost model as a design choice. The underlying assumption behind this choice is the naive modernist assumption that through the time the world is progressing on average. The shapley values indicate this modernist assumption. The red dots representing the latest years are residing on the negative scale which means a drop in NEET values.

4 Results

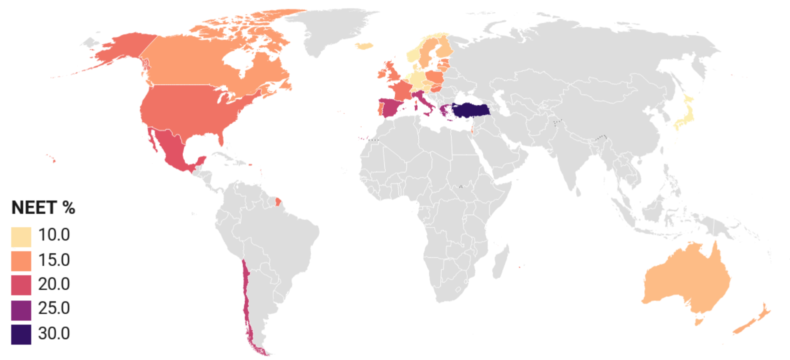

Figure 4: Map of Forecasted NEET% in 2020.

In figure 4 The countries colored in bright yellow are the ones forecasted to keep their NEET rates under 10%. Those countries are the Netherlands, Japan, Norway, Germany, Denmark, and Switzerland. Austria and Luxembourg are also forecasted to have NEET levels which are slightly above 10% mark (10.03% each). 20 OECD countries remain in the 10%to 20%band. The worst performing countries are Chile, Spain, Italy, Greece and Turkey. The forecasted NEET rates for the first 4 of the low performers float in the 20% to 24% mark. Turkey, as always being the worst performer of all OECD countries in the NEET rates, is again the worst performing country with 30.3% forecasted NEET rate.

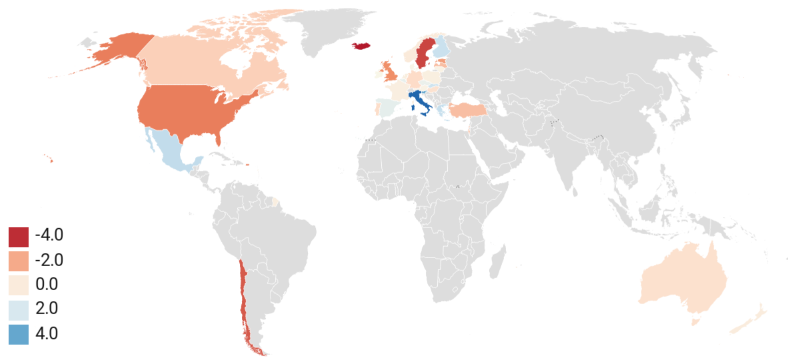

Figure 5: Map of Difference in NEET% between 2015 and 2020

The difference is calculated by subtracting 2020 values from the 2015 values. Negative values indicate a surge in NEET values whereas positive values indicate decrease in NEET. There are 17 countries where NEET values are forecasted to be less in 2020 in reference to their NEET values in 2015. For the remaining 16 countries the contrary is true. Among the progressing countries only Slovenia and Italy are expected to have major improvements (Slovenia - 3.99%, Italy, 5.61%). For the other 15 progressing countries the progress rate is between +0% and 2.5%.

On the worsening side of the picture, there are 16 countries which are forecasted to fall back from 2015 NEET rates in 2020. Iceland, Sweden, Chile, United States of America, United Kingdom, and Estonia are the countries which are forecasted to have NEET increase more than 2%.

5 Discussion

As figure 6 shows, the mean NEET% in OECD region is forecasted to increase sharply, 1.7%, in 2020. The mean NEET% was decreasing almost steadily between 1997 and 2008. It took 10 years for OECD countries to bring the NEET rates under the 2008 value. The current coronavirus pandemie brought another global economic recession. Unlike the 2008-2009 crisis, there is a different recovery path on the table, the "L" shaped recovery. This pessimistic recovery path projects the globe to remain in the crisis for an extended time period. If it becomes the case, then the NEET rates are also expected to remain high for a decade.

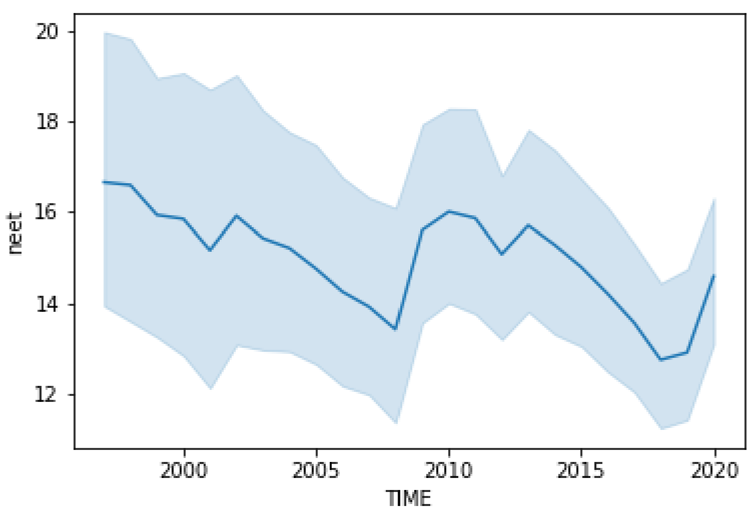

Figure 6: Mean NEET% in OECD area between 1997 and 2020

The last danger lurking around for the NEET is artificial intelligence led automation. If the shift is not managed well by the governments with a multifaceted agenda, then the A.I based automation can narrow down the number of jobs which were traditionally taken by the NEETs. The hard task for the policy makers around the globe seems to be finding a balanced pathway which simultaneously leverages AI based automation to help the economies bounce back; while not creating a welfare dependent masses of youth who are caught unprepared automatization era.

References

Bradley, S., Migali, G., & Navarro Paniagua, M. (2019). Spatial variations and clustering in the rates of youth unemployment and neet: A comparative analysis of italy, spain and the uk. *Journal of Regional Science*.

Breiman, L., et al. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science*, *16*(3), 199–231.

Bruno, G. S., Marelli, E., & Signorelli, M. (2014). The rise of neet and youth unemployment in eu regions after the crisis. *Comparative Economic Studies*, *56*(4), 592–615.

Caroleo, F. E., Rocca, A., Mazzocchi, P., & Quintano, C. (2020). Being neet in europe before and after the economic crisis: an analysis of the micro and macro determinants. *Social Indicators Research*, 1–34.

Coles, B., Godfrey, C., Keung, A., Parrott, S., & Bradshaw, J. (2010). Estimating the life-time cost of neet: 16–18 year olds not in education, employment or training. *York: University of York*.

Coles, S. (2017). Social enterprise and the well-being of young people not in education, employment or training. *Social Determinants of Health: An Interdisciplinary Approach to Social Inequality and Wellbeing*, 181.

Denny, S., Hazenberg, R., & Seddon, F. A. (2010). Social enterprises and neets: ethical and effective entrepreneurial skills programmes?

Dietrich, H. (2013). Youth unemployment in the period 2001–2010 and the european crisis–looking at the empirical evidence. *Transfer: European Review of Labour and Research*, *19*(3), 305–324.

Friedman, M. (1977). Nobel lecture: inflation and unemployment. *Journal of political economy*, *85*(3), 451–472.

Holte, B. H. (2018). Counting and meeting neet young people: Methodology, perspective and meaning in research on marginalized youth. *Young*, *26*(1), 1–16.

I.M.F. (2020a). *Inflation rate, average consumer prices annual percent change (indicator.* (`https://www.imf.org/external/datamapper/PCPIPCH@WEO/OEMDC/ADVEC/WEOWORLD?year=2020`, Last accessed on 30 July 2020)

I.M.F. (2020b). *Unemployment rate percent(indicator.* (`https://www.imf.org/external/datamapper/LUR@WEO/OEMDC/ADVEC/WEOWORLD?year=2020`, Last accessed on 30 July 2020)

Kelly, E., & McGuinness, S. (2013). *The impact of the recession on the structure and labour market success of young neet individuals in ireland* (Tech. Rep.). ESRI Working Paper.

Krause, K., & Liebig, T. (2011). The labour market integration of immigrants and their children in austria.

Liebig, T., Kohls, S., & Krause, K. (2012). The labour market integration of immigrants and their children in switzerland.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., ... Lee, S.-I. (2020). From local explanations to global understanding with explainable ai for trees. *Nature machine intelligence*, *2*(1), 2522–5839.

MacMillan, L., Gregg, P., & Britton, J. (2012). *Youth unemployment: The crisis we cannot afford* (Tech. Rep.). Technical report, The ACEVO Commission on Youth Development.

O.E.C.D. (2020a). *Real gdp forecast (indicator.* (Accessed on 30 July 2020) doi: 10.1787/1f84150b-en

O.E.C.D. (2020b). *Trade in goods and services forecast (indicator.* (Accessed on 30 July 2020) doi: 10.1787/0529d9fc-en

O.E.C.D. (2020c). *Youth not in employment, education or training (neet) (indicator.* (Accessed on 30 July 2020) doi: 10.1787/72d1033a-en

Phillips, A. W. (1958). The relation between unemployment and the rate of change of money wage rates in the united kingdom, 1861-1957. *economica*, *25*(100), 283–299.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). Catboost: unbiased boosting with categorical features. In *Advances in neural information processing systems* (pp. 6638–6648).

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Model-agnostic interpretability of machine learning. *arXiv preprint arXiv:1606.05386*.

Rodriguez-Modroño, P. (2019). Youth unemployment, neets and structural inequality in spain. *International Journal of Manpower*.

Rodwell, L., Romaniuk, H., Nilsen, W., Carlin, J. B., Lee, K., & Patton, G. C. (2018). Adolescent mental health and behavioural predictors of being neet: a prospective study of young adults not in employment, education, or training. *Psychological medicine*, *48*(5), 861–871.

Ryan, P. (2001). The school-to-work transition: a cross-national perspective. *Journal of economic literature*, *39*(1), 34–92.

Saabas, A. (n.d.). Treeinterpreter library (2019). *URL https://github.com/andosa/treeinterpreter*.

Strand, S. (2007). *Minority ethnic pupils in the longitudinal study of young people in england. dcsf research report rr-002.* Department for Children, Schools and Families.

Unit, G. B. C. O. S. E. (1999). Bridging the gap: new opportunities for 16-18 year olds not in education, employment or training.

Williamson, H. (1997). Status zer0 youth and the. *Youth, the'underclass' and Social Exclusion*, 70.