Skill Analytics for Personal Growth and Success

Anis Hazirah Mohamad Sabry
Faculty of Computing and Informatics
Multimedia University
Cyberjaya, Malaysia
1211300373@student.mmu.edu.my

***Abstract*— Asia is renowned for its remarkable economic achievements, driven by the diligent labour of its workforce and rooted in a complex cultural fabric. However, there is a lack of study that examines the reasons behind the reputation of Asia's work culture for its diligent employees and high production. The objective of this study, with the help of the Programme for the International Assessment of Adult Competencies (PIAAC) dataset, is to evaluate the factors that impact working hours, analyse the discrepancy between qualifications and employment, and explore the strategies individuals adopt to address skill gaps with the use of machine learning. The aim of these aims is to shed light on important factors that affect work productivity in Asia, thereby enhancing our comprehension of the region's economic viability and expansion.**

Keywords—Asia, PIAAC, working hours, qualification discrepancy, skill gap

I. INTRODUCTION

A. Overview

In this era of technological advancements, countries across the globe are in a rat race to be able to maximise the possible avenues of their country as efficiently as possible. For this to happen, mass amounts of data concerning the population is required to uncover and disseminate information The Organisation for Economic Growth and Development (OECD) took the initiative to make this a possibility with the implementation of a comparative survey of adults titled 'The Survey of Adult Skills'. With this they had developed the Programme for the International Assessment of Adult Competencies (PIAAC), an international level survey that divulges into the critical need about the distribution of knowledge, skills, and characteristics that have become vital key factors for an optimised participation in modern societies. The PIAAC data aims to provide policymakers a baseline of their populations' levels of knowledge, skills, and competencies as elements towards personal and societal success, creating a gauge of economic outcomes and recommendations towards enhancing the countries' human capital. They provide a deeper understanding of the processes behind skill gains, loss, and retainment, allowing a more dynamic knowledge base in regards to these issues.

In recognising the scarcity of research done on Asian countries, particularly those that encompass the region as whole, this project aims to dive deeper towards understanding the sustainability of Asia's economy. Particularly in regards to the role of adaptive problem solving as a key to personal growth and success in the context of the Asian working environment. Through an analysis of the relationship between education levels, working environments, and problem-solving skills, the research aims to illuminate the key features influencing work efficiency in the region.

B. Problem Statement

The motivation behind embarking on this task to unravel and understand the intricacies inherent in Asian working culture stems from the necessity to comprehend the profound significance of work ethics in this region. By delving into the driving forces behind Asian work culture, we seek to uncover patterns and insights crucial for addressing challenges related to employment, skill utilisation, and the educational landscape of the workforce.

C. Research Objectives

1. Investigating why people work more or fewer hours than required.
2. Conducting association rule analysis to identify factors influencing the mismatch between highest academic qualification and job requirements.
3. Investigating how individuals address the skill gap when they lack the necessary qualifications for their current profession.

D. Project Scope

The study focuses on analysing occupational mismatch, with the focus of problem solving skills, using the first cycle (2011 to 2017) of PIAAC dataset for Asian countries. Those countries are Japan, South Korea, Kazakhstan, and Singapore. This project intends to:

1. Examine the factors and relationships concerning work attributes. This involves researching variables that would influence working over or under the average working hours.
2. Identifying if a skill mismatch impacts working hours and comparing the information across different work sectors.
3. Classifying the worker mismatch by their education mismatch and working hours.

II. LITERATURE REVIEW

A. International Analysis of Problem Solving Skills in the Context of Work

Education has a high correlation with a high level of problem solving. [1] In regards to work variables, it was observed that individuals who were employed and received payment, had exposure to relevant experiences, regularly dealt with problem solving tasks, possessed advanced computer skills, worked in more specialised occupations, and earned a higher monthly income were more likely to achieve a high score in terms of problem solving.

With this in mind, it should be noted that using this as a means of measurement may not be equitable for individuals from diverse backgrounds. Immigrants, particularly in the United States, frequently face racial prejudice. Asian Americans and black workers often possess credentials that far exceed the minimum requirements for their respective occupational categories. This phenomenon is particularly prevalent among first-generation immigrants in comparison to third-generation immigrants. [2]

Similarly, Vocational Education and Training (VET) workers face distinct and ever-changing challenges, which differ from those encountered by individuals pursuing general education courses. These workers are prone to having lower problem-solving skill scores and often require additional support compared to others in order to enhance their professional competencies. The reason for this is the discrepancy between VET courses and general education courses. [3]

On a similar note, WPL is an essential component of lifelong learning (LLL) since the advancement of new technologies necessitates a greater level of expertise in applying skills to adapt to them. Out of the 60 articles found, only 7 of them had relevance towards WPL. These topics range from VET, opportunities to informal learning, and participation in learning at work. The authors note the scarcity of research based on WPL and advocate for its untapped potential. According to them, the PIAAC data has the potential to offer a more thorough examination of WPL, including interesting factors like learner profiles and behaviours, to gain a deep understanding of how an individual implements WPL. [4]

Further complexities arise once the occupational sector is considered. Many educators frequently demonstrate a skill mismatch, which is characterised by a discrepancy between the demands of their current job and their highest level of education. The misalignment leads to a perception of educators having reduced levels of professionalisation. [5]

Another perspective to consider is investigation of the PSTRE via process data. [6] A multiclass hierarchal classification was conducted with the models Random Forest Classifier and Support Vector Machine (SVM). Their study concludes that hierarchical classification models moderately outperform flat classification models in predicting proficiency levels. With the emergence of Artificial Intelligence (AI) tools being widespread, the authors heavily emphasise the use of process data as a validity check to see if the response was entirely human.

Similarly, another research [7] uses a top-down approach, creating rule-based indicators for information processing methods. The authors utilised the log data for countries where the PSTRE was taken, and analysed behaviours during the simulated web search environment, showcasing the frequency and consistency of these behaviors and their impact on task success. Patterns of success were most commonly indicated by variables such as age, gender, education, and reading and evaluation skills. It highlights the effectiveness of life-long learning, and emphasises on the concerns of adult digital competencies in this current age.

B. Analysis of PIAAC Data for Asian Regions

It can be argued that the PIAAC data is not a proper representation of the human capital found in Asian regions. There is a lack of consideration towards the cultural differences in OECD'S reports. [8] OECD does not take into the account the historical attributes that would result in the economic prosperity that blooms from the years of detrimental conditions these countries had the unfortunate luck of experiencing. They commented that policy recommendations are simplistic in its views and too ideological, as opposed to being rooted in reality.

The authors also observed that education is not the primary determinant of economic growth, but rather a contributing factor. They propose that policymakers should not prioritise education as the primary basis for development. Instead, they advocate for a more moderate approach in devising solutions to achieve their goals.

C. Analysis of Problem Solving Skills Asia in the Context of Work

To fully understand the roles at play to output the best possible productivity levels in a work environment, a study [9] builds upon previous research of examining age effects on productivity, the effect of ICT skills on productivity, followed by assessing how job training participation improves productivity. They find that as workers get older, their productive levels and skills may only be maintained with adequate participation in job training.

Another study [10] further explore this subject by categorising older workers based on typology. The authors examine the characteristics of older workers by analysing their engagement in learning activities, which encompass their readiness to learn, their involvement in informal learning inside the workplace, and their participation in informal learning opportunities. Their research uncovers three distinct categories: The Dormant Workers with Low Skills, The Educated Workers in the Public Sector, and The Educated Workers in Flexible Working Conditions. They emphasise the need for organisations to embrace a non-traditional approach towards ageing workers, allowing flexibility in order for them to fully satisfy their needs and job objectives. Consequently, the support given to these workers aids in maintaining a competent level of problem solving skill.

In another study, [11] reported consistent findings regarding varying levels of problem-solving competencies across different age cohorts. Typically, older workers tend to exhibit a decrease in their problem solving abilities, indicating a strong correlation between age and experience. Another argument for this reasoning is that adolescents are exposed to a significantly greater number of problem-solving opportunities in their educational experiences. Furthermore, the rapid introduction of new technologies poses challenges for older generations who struggle to comprehend them as swiftly as their younger counterparts. The authors demonstrated that engaging in interactions with colleagues within the workplace is able to counteract this effect.

In addition, another research [9] propose another way to address the degradation of problem solving competencies in older generations. They vouch for conduction of vocational training as a supplement in preparing workers with adequate skills. This fosters the workers' understanding of the latest technological advancements, ensuring they stay informed and enhancing their career opportunities by incorporating life-long learning.

Additionally, a study [12] claims skill usage is the most effective way to avoid deterioration of skill competencies. They demonstrate that the utilisation of skills has a beneficial effect in both professional and domestic settings. Moreover,

their research showcases how problem solving skills are often associated with basic background information, cultural capital defined by the number of books kept at home, and skill usage whether it is in a personal setting or a work setting.

In looking at the issue of skills and job mismatches in the labour market within the context of an Asian region. A study [13] reveals that the influence of education on work satisfaction is relatively smaller compared to the significance of workplace skills. On the other hand, the combination of these two factors leads to incomes that are lower. In the case of over qualified workers however, there is a reduction in productivity the longer they are employed, most likely due to the lack of opportunities available to enhance their skillset. The results of this suggest that those that are under qualified may eventually progress into the modal average of their occupation overtime. It is important to note however that these factors vary depending on the sector of occupation.

Another study [14] examine the likelihood of being employed in Japan and South Korea. The authors analysed individuals aged 25 to 55, irrespective of their employment situation. Their research demonstrates that an increase in years of formal education and informal learning has a substantial effect on the likelihood of employment for people in both nations, however the influence is more pronounced in Japan compared to South Korea. Nonetheless, they regarded a comprehensive focus on enhancing skill development as a more advantageous investment for workers, highlighting that their education and training systems have room for improvement to be able to keep up with the demands of their current and future needs for skill usage.

While there exist studies concerning the PSTRE data, working environments, and named Asian countries, there is a lack of research done specifically on the Asian region as a whole and its' working environment. This study aims to fill in this research gap by addressing the working hours and skill mismatch found in workforce environments of Asian countries.

III. THEORETICAL FRAMEWORK

A. The PIAAC Data

The PIAAC data collection comprises three assessment cycles. As of writing, the PIAAC has completed the first cycle of data collection which took place between 2011 to 2017. The second cycle is currently underway and is expected to take place between 2024 to 2029. The survey evaluates participants aged 16 to 65 in various countries through either traditional pen and paper methods or digitally using a computer. [15]

In this research, the countries that were analysed are the four Asian countries that are available in the dataset. The countries mentioned are South Korea, Singapore, Japan, and Kazakhstan. Although Indonesia is an Asian country, their method of data collection was only via the use of pen and paper due to most of the population not being familiar with the use of a computer and thus, did not partake in the PSTRE assessment. Despite Turkey and Israel being geographically located within West Asia, they are considered Middle East. Alongside that, the current ongoing political imbalance between those countries may affect the results of this study and thus, are not included.

The combined number of participants from these countries amounted to 23,463. It should be noted that the Singapore dataset does not have an adequate amount of data for the ages of its participants. Hence, the variable of age will not be taken into account in this study.

B. Questionnaire

The PIAAC data is an extensive repository of information, offering a diverse range of inquiries to enhance comprehension of an individual. The features utilised for this study were obtained from another research [3] where the authors categorised the available features into four distinct subgroups. Those subgroups are demographic characteristics, work and education, work skill use and learning, and everyday life skill use and learning. These subgroups are further refined to only include features that consider problem solving or ICT questionnaires. This aids in generating a more complete picture of the participants, adding on more possibilities to comprehend and analyse skill sets within the context of other factors that could possibly influence it.

In order to maintain the dataset's relevance to our research, we have excluded participants who were unemployed from the dataset. In addition, the PIAAC data is presented in an encoded format. Therefore, the codebook was utilised to decode each value, ensuring that the data is easily readable and maximally effective for this research.

C. Skill Domains

The OECD asserts that the PIAAC assessment examines three cognitive skills domains: literacy, numeracy, and PSTRE. Literacy is a cognitive ability that involves understanding and interpreting written information in various forms and styles, and being able to respond to it appropriately. Numeracy is a skill that involves not just the ability to solve mathematical problems, but to contend with an array of possible representations of numerical issues, not limited to merely arithmetic knowledge and computation. [15]

As for PSTRE, this domain primarily revolves around a specific class of problems dealing with ICT tools. Often considered as 'computer literacy', this skill encompasses the capacity of an adults' use with these tools and applications, as well as the ability to maintain and handle technology-rich environments. Having said that, the main objective of analysing this skill is to assess the capacity of accessibility, process, evaluation and analysis capabilities of adults within the realm of ICT. The cognitive dimensions used to assess PSTRE include goal setting, monitoring progress, planning, accessing and evaluating information, and selecting, organising, and transforming information. Environments such as web, spreadsheet, and email environments were used to influence the difficulty in performing each of these tasks.

In the case of literacy and numeracy, all participants successfully completed both assessments. For PSTRE however, participants who do not have access to or have no experience with using a computer were unable to participate in this assessment.

TABLE I. DEFINITION OF EACH PIAAC SKILL DOMAIN

Skill	***Definition***
Literacy	Understanding, evaluating, using and engaging with written texts to participate in society, to achieve one's goals, and to develop one's knowledge and potential
Numeracy	The ability to access, use, interpret and communicate mathematical information and ideas, in order to engage in and manage the mathematical demands of a range of situations in adult life
Problem Solving in Technology Rich Environment (PSTRE)	The ability to solve problems for personal, work and civic purposes by setting up appropriate goals and plans, and accessing and making use of information through computers and computer networks

PSTRE can be divided into four proficiency levels that determine the knowledge and skills needed to accomplish tasks at each level. [3] The levels are derived from the problem solving scale score using plausible values, indicated as PVPSL1 to PVPSL10. The lowest level includes scores from 241 to 290 points, the second lowest includes scores from 291 to 340, and the highest level includes a scoring that is from 341 and above. Each of these are labelled as 'weak performers', 'moderate performers', and 'strong performers' respectively. Those that obtain a scoring below the first level are considered 'at risk'.

To optimise the dataset's performance to answer the problems in this research, the dataset is refined to include only those that have their problem-solving skills variable filled out. This is because no imputation will be involved to keep the data as accurate as possible. Those who do not have problem solving skill scores are participants who do not have access to computers or do not have any experience with modern technology and thus are not relevant in our research. Along with these features, the tenth problem solving variable, PVPSL10, will be used. This variable is sufficient to represent all 10 problem solving variables. [3] An additional variable that dictates the performance level of the participant based on their PVPSL10 score will also be included.

D. Work Environment of Asian Countries

1) Skill Mismatch

Skill mismatch refers to the type of imbalance between the skills possessed by a worker and the skills required to perform the demands of their current job. The evaluation method encompasses a wide range of discrepancies, encompassing both qualitative and quantitative aspects, as well as formal and informal education. This imbalance forms challenges at different stages of a worker's career. [13] Consequently, the labour force is affected by a decrease in the number of workers who are no longer able to meet the requirements of their jobs.

The PIAAC data uses the International Standard Classification of Education (ISCED) 1997 to label each of the participants' education levels. ISCED level ranges from 1 to 6, with subcategories of A, B, or C. As each country has its own unique education system, each of them has been converted into its appropriate ISCED value to ensure uniformity within the PIAAC dataset. [15]

In order to conduct our research, we will categorise the skill mismatch based on the definition used by another study [2]. An overmatch is defined as possessing credentials that exceed the average expected credentials in their occupational sector, whereas an undermatch is defined as possessing credentials lower than the expected average of their occupational sector. Individuals whose educational qualifications align with the job requirements will be regarded as 'equal. The study will take the individuals' highest level of education and the qualifications necessary for their job to define them. Those with a foreign level of education will be converted to the relevant ISCED levels.

2) Working Hours Across Asian Countries

This research will consider the relevant legislation on working hours in each country to accurately determine the appropriate working hours. Normal working hours refer to the standard hours of work for an employee, excluding any instances where the employee has agreed to work additional hours with their employer. The maximum possible working hours will be considered as working hours that include those factors.

For this research, the range that will be used as the average working hours will differentiate between countries. In Japan, the average range falls between 40 and 42. In the case of Singapore, the average range will be between 44 and 48. Both South Korea and Kazakhstan have the same average range of 40 to 52. Individuals who work less hours per week than the national average will be categorised as 'below average', while those who work more hours per week than the national average would be categorised as 'above average'. The table below details the normal and maximum working hours for each country.

TABLE II. TABULATED NORMAL AND MAXIMUM POSSIBLE WORKING HOURS PER WEEK

Country	**Japan**	**South Korea**	**Kazakhstan**	**Singapore**
Normal working hours per week	40	40	40	44
Maximum possible hours per week	42	52	52	48

E. Machine Learning Models

Three machine learning models will be used and finetuned for the purpose of this project. Those models are random forest classifier, decision tree classifier, and SVM.

Each of these models will be finetuned to achieve the best possible outcome for this project.

The random forest classifier is an algorithm that combines multiple decision trees into a singular ensemble. It grows each decision tree by feeding it random sampling subsets. Any data not included is used for validation. As each tree grows, the nodes splits into a subset of predictor variables and stops when a predefined number of leaf nodes or impurity threshold has been achieved. The output is determined by aggregating the results from each individual tree, thus improving the accuracy and reducing the chances of overfitting compared to using a singular decision tree classifier. [6]

The next model is a decision tree. A decision tree is as the name implies, a tree of which consists of branches, nodes, and leaves that aid in both regression and classification tasks. It is the tree that makes up the ensemble model, random forest. Similar to the random forest model, it has three criterions for splitting. Namely the gini impurity which measures the likelihood of incorrect classification of a randomly chosen element from the data, entropy which measures the amount of uncertainty of nodes, and log loss or logarithmic loss which measures the uncertainty of predictions. While both entropy and log loss measure similar things, entropy guides the decision tree for purer subsets, while the log loss penalizes incorrect predictions for a better overall probability prediction.

The last machine learning model is an SVM. In a SVM, kernels are considered optimal boundaries that classify the dataset in different regions. A kernel's function is to transform the input data into its' required form. Some of these kernels include linear kernels such as the linear kernel and nonlinear kernels such as the polynomial kernel, Gaussian radial basis function (RBF) kernel, and the sigmoid kernel. [6]

F. Association Rule Analysis

Colloquially called 'market basket analysis', association rule mining will be implemented in this research to uncover factors of work mismatch between a workers' highest qualification and work qualification. Association rule mining functions by sorting item sets within a dataset and unveiling associations between each distinct element. Key metrics of support, confidence, and lift are used to define and identify these associations.

The associations produced are referred to as 'rules', which consist of itemsets representing both the antecedent (events that occur) and the consequent (outcome). A support is defined as the frequency of an itemset occurring, while confidence is defined as the probability of a consequent having existing antecedents. Finally, there is lift. Mathematically, it is defined as the likelihood ratio of having an additional element given the presence of another element that may be associated with it. Lifts can introduce the strength of an association between two elements. It refines the association rules even further by preventing misinterpretation of rules with a high confidence metric but low association. Lifts with a value greater than 1 indicate a positive association, while lesser than 1 indicates a negative association. [16]

The association rule mining is applied on to the dataset to find common factors that are associated with a mismatch in educational qualification and work requirements, alongside with identifying how under qualified workers can mitigate and retain their current employment.

IV. RESEARCH METHODOLOGY

Deciphering the dataset and finding the patterns to solve the research objectives involves understanding the codebook provided by the OECD. In the codebook, there are columns that denote the sequence of the item in the dataset, the name of the item, the label which dictates the assessment question, the type of data (be it an integer or a string), the level which can be nominal, ratio, scale, or ordinal, the missing scheme indicator, and a link to a separate spreadsheet that entails all the true meanings behind each encoded answer.

A. Preliminary Exploratory Analysis of the PIAAC Data

A preliminary exploration of the dataset was conducted to have an initial look at the data that will be further analysed. The histogram displays that for the country of Kazakhstan, the count of age peaks at around 30, while the age in Japan peaks near the 40th mark. For Korea however, several peaks in ages are found in the dataset, those peaks can be found near 20th, 30th, and the highest being by the 40th mark in age.

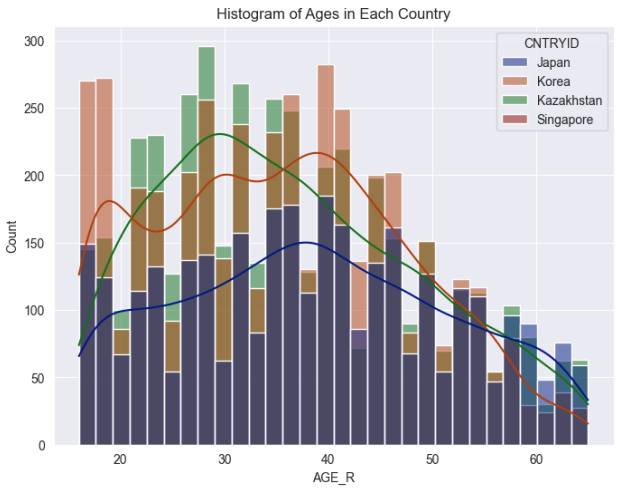

Fig. 1. Histogram of ages in each country

Seeing as how the age varies across the dataset, a closer inspection of it against the hours of informal learning showcases how most of the participants are between 0 to 500 hours of informal learning regardless of age. This case is most true for participants from Kazakhstan. For the case of Korea however, the informal learning hours vary regardless of age, with some having almost 2000, while others are spread across 0 to 2000 informal learning hours. In Japan, the informal learning hours are not as high as other countries, with a

smaller amount being above 500 hours when compared to Korea and Kazakhstan.

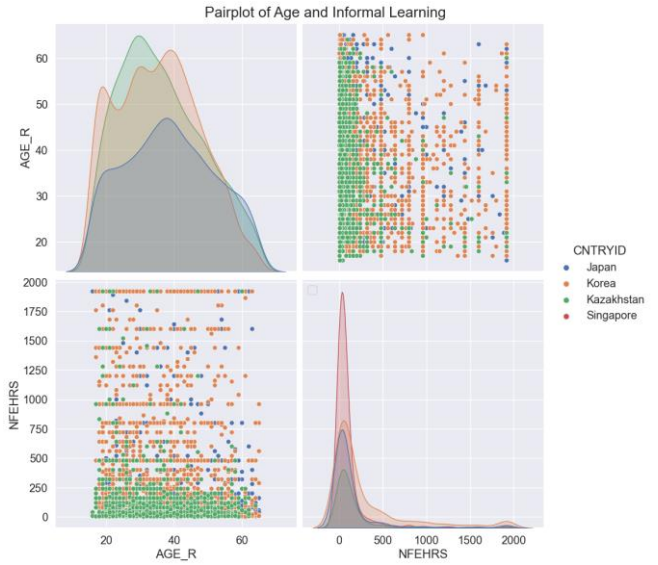

Fig. 2. Pairplot of age and informal learning

A countplot of the participants' subjective status of their current work or work history shows that a majority of them are full time workers, with 1943 from Japan, 2489 from South Korea, 2607 from Kazakhstan, and 2612 from Singapore. Meanwhile those who are part-time workers have a count of 490, 596, 560, and 204 from Japan, South Korea, Kazakhstan, and Singapore respectively. The dataset also features students, having 389 from Japan, 747 from South Korea, 394 from Kazakhstan, and 606 from Singapore. Only 3 participants in total did not state their current work status. Other variables include those who are fulfilling domestic tasks or looking after family members with a total count of 1679, 642 unemployed participants, 320 in retirement or early retirement, an undefined other category with a count 253, interns with a count of 69, permanently disabled participants with a count of 48, and those in compulsory military or community service with a count of 124.

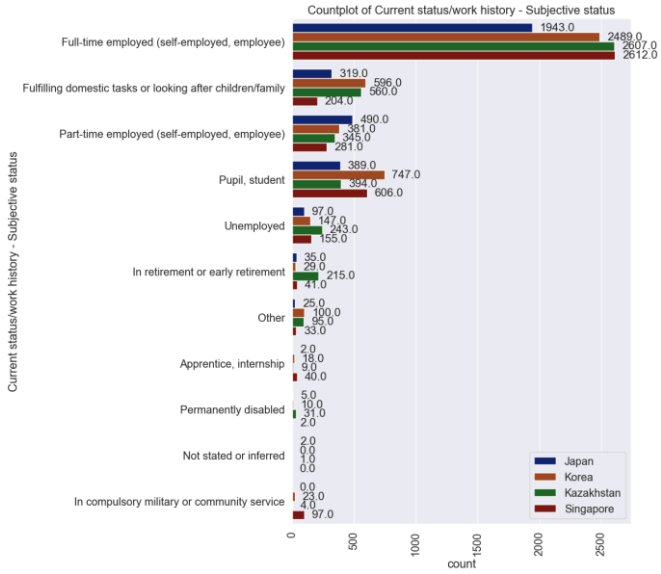

Fig. 3. Countplot of subjective status for current work status/history

B. *Age and Working Hours*

To illustrate the normal working hours, we analyse the participants' working hours and categorise them by nation. Boxplots are defined by their quartiles, which are represented by the median, first quartile (q1), and third quartile (q3). In Japan, the median value is 40, with a q1 of 35 and a q3 of 50. The median in South Korea is 44, with a q1 of 40 and a q3 of 50. Kazakhstan's median is 40, which is equivalent to the q1, while the q3 is 48. The median value in Singapore is 44, with the q1 at 40 and the q3 at 50. Several outliers for working hours can be seen, particularly those that exceed 100 working hours per week can be found in Japan and Singapore.

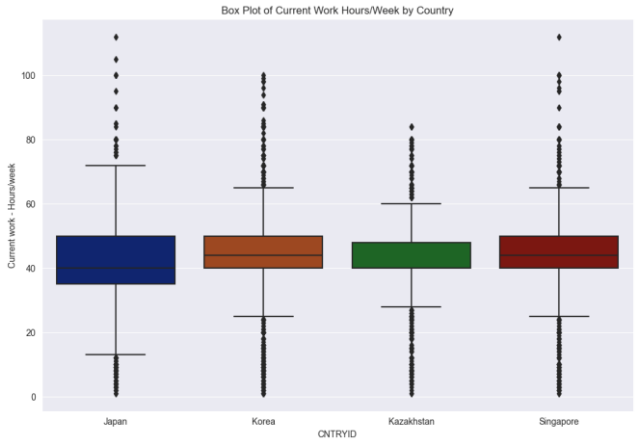

Fig. 4. Boxplot of current work hours per week by each country

The average working hours in Japan varies across the ages of below 20 up to the age of 65. According to the PIAAC, we can see that the average number of working hours consistently rises until an individual reaches between the age range of 20 to 30. From that point onwards, the average number of hours worked fluctuates between about 50 hours per week and fewer than 40 hours per week until the age of 50. From this point, the trend begins to decline, with a sudden increase in average working hours beyond the age of 60 before ultimately declining.

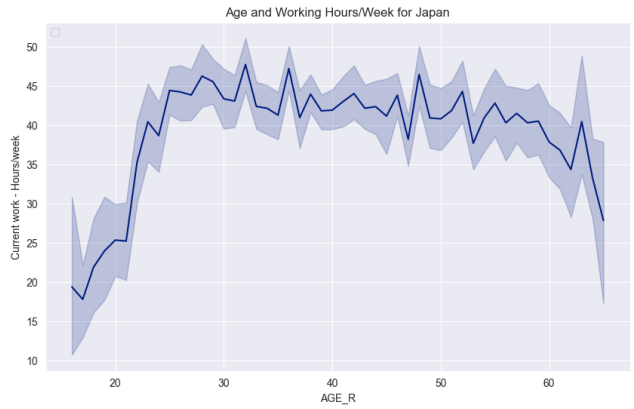

Fig. 5. Lineplot of age and working hours per week in Japan

In South Korea, the number of working hours gradually rises until an individual reaches the age of 20. Subsequently, the weekly working hours increase before continuously remaining between 40 and 50 hours each week.

After reaching the age of 50, there is a noticeable decline in working hours, followed by two periods of increased working hours after the age approaches 60. Those increases are 50 hours and 60 hours. Subsequently, the trend declines to nearly an average of 30 hours before subsequently rebounding.

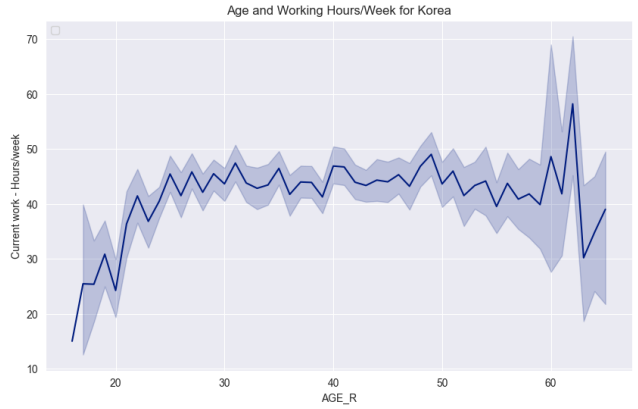

Fig. 6. Lineplot of age and working hours per week in South Korea

At the age of 20, the working hours in Kazakhstan rise to approximately 40 hours each week. The pattern remains constant within the range of 40 to 50 hours from the ages of 20 to 60. After reaching the age of 60, there is a noticeable and significant decline in the trend, with the minimum number of working hours per week dropping to about 30. Multiple upward spikes are shown before ultimately rising.

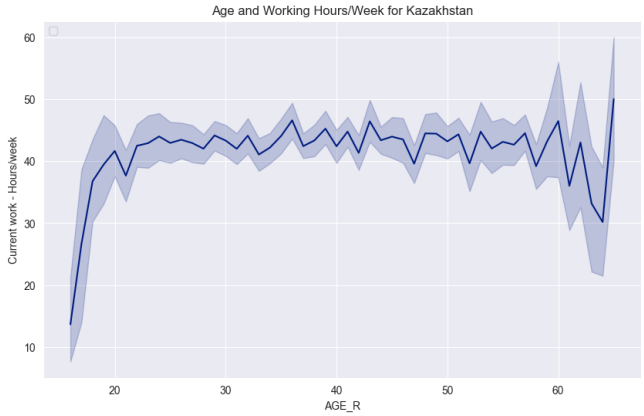

Fig. 7. Lineplot of age and working hours per week in Kazakhstan

C. Skill Use and PVPSL Performance

The pairplot of index of skill use at home and informal learning for non-job related reasons reveals that the majority of the index of skill use falls within the range of 0 to 6. Informal learning hours in countries typically range from 0 to 500, while the skill use index ranges from 0 to 4. The remaining data remained sparsely distributed up until 1500 hours, with just a small portion falling above that threshold until 2000.

Fig. 8. Pairplot of skill use at home and informal learning for non-job related reasons

The pairplot of index skill use at home and at work showcases a linear relationship between the two features, displaying a linear line that entails the skill use at home and at work are correlated to one another.

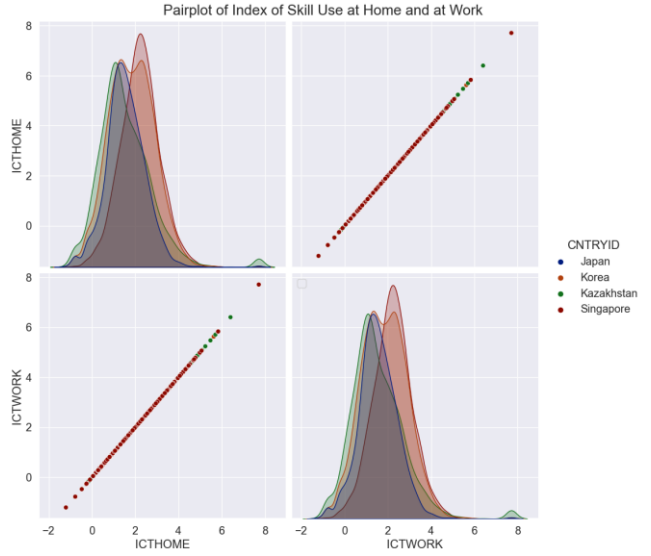

Fig. 9. Pairplot of skill use index at home and at work

The problem solving performance in all countries indicates that the number of individuals with strong performance is below 500 in each country. Japan has the highest count in this category with a total of 469, followed by Singapore with 423. South Korea has a count of 231, while Kazakhstan has the lowest count which is 52. The number of moderate performers individuals from South Korea and Singapore surpasses 1500, with South Korea having a count of 1741 and Singapore with 1648. Japan is currently sporting a count of 1414, while Kazakhstan has the lowest count which is 883. Kazakhstan has the greatest count of weak performers, with a count of 2446, while South Korea has almost 1922. Singapore and Japan fall within the range of 1000 to 1500, with Singapore having 1356 and Japan having 1065. The category labelled as 'at risk' is regarded as the class with the poorest performance. Kazakhstan holds the largest

number in this category, surpassing 1000 with 1123 participants. South Korea and Singapore have numbers of 646 and 644 respectively. On the other hand, Japan has the lowest number of performers considered 'at risk', with a total of 359.

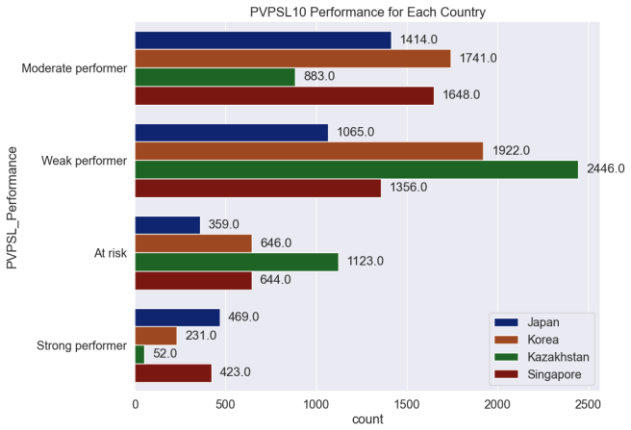

Fig. 10. Countplot of PVPSL performance for each country

The histogram of performance scores reveals that a significant proportion of individuals with moderate performance from all four Asian countries tend to cluster around the scoring range of 200 to 300. Weak performers on the other hand in Japan and South Korea have a high number of participants scoring between the range of 280 to 290, while Kazakhstan has the highest peak between the scoring range of 260 to 270. Singapore on the other hand, has most of its weak performers scoring at around 270. For those who are at risk, the histogram appears left-skewed, with the participants from this category mostly comprising those scoring between 220 and 240. Strong performers on the contrary exhibit a right-skewed distribution, with the majority of scores falling within the range of 340 to 360.

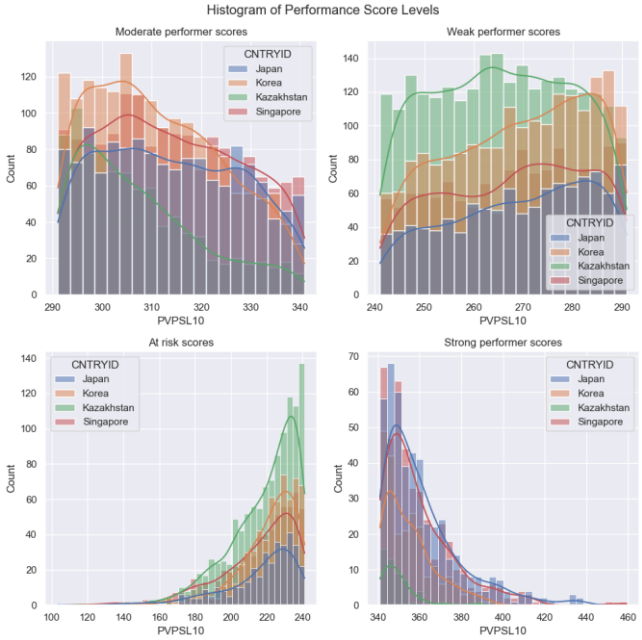

Fig. 11. Histogram of PVPSL performance scores by levels

D. Education and Qualification Mismatch

A look at the count of highest education levels in the dataset reveals that most of the participants have a level of ISCED 5A, bachelor's degree, with a count of 980 from Japan, 1151 from South Korea, 1077 from Singapore, and the largest count belongs to Kazakhstan with 1616. Besides that, another notable value is ISCED 5B with 696 from Japan, 903 from South Korea, 273 from Kazakhstan, and 952 from Singapore. Another value that has a high count is ISCED 3A-B, where Japan has 813, South Korea has 1101, and Kazakhstan has 645. For this value, there are none from Singapore. A possibility for this reason is that it's between their lower secondary education (ISCED 2) and upper secondary education (ISCED 3 (without distinction A-B-C, of 2 years or more)) and thus, may not have an equivalent in the ISCED system. ISCED 3 (without distinction A-B-C, of 2 years or more) shows a high count for those in Kazakhstan and Singapore, with counts of 895 and 888 respectively.

Other values include ISCED 2 with a total of 1271, ISCED 3C of 2 years and more with a total of 986, ISCED 3C of shorter than 2 years with a count of 64, ISCED 4 (without distinction A-B-C) with a total of 59, ISCED 5A, master degree, with a total of 655, a total of 24 foreign qualifications, ISCED 6 with a total of 79, ISCED 1 with a total of 130, and ISCED 4A-B with a total of 1099. The lowest level, that is below ISCED 1 or no formal qualification, has a total of 44, with the majority being from Kazakhstan. A total count of 5 participants did not state their highest education level.

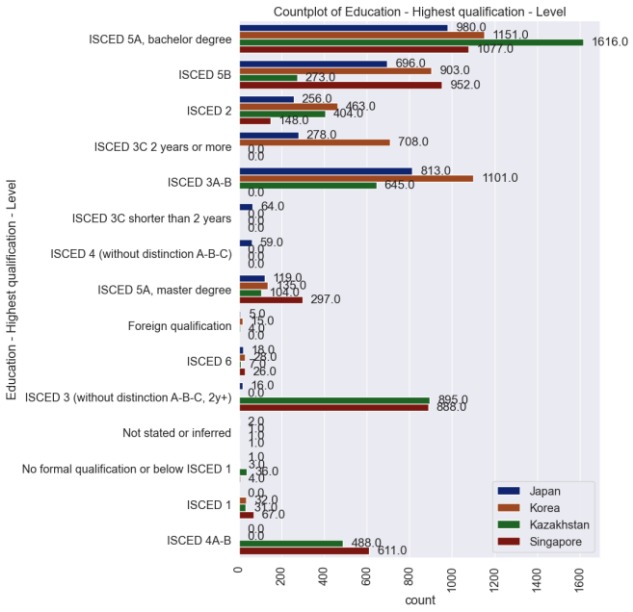

Fig. 12. Countplot of highest education level

In looking at current work requirements, most of the participants do not state their current work requirements. Most jobs require an education of an ISCED 5A, bachelor degree, with a count of 2971 participants' jobs having this work requirement. Besides that, 1962 of the participants' job requirements is ISCED 5B and 1532 participants' jobs require an education of ISCED 3A-B. The other educational requirements have a count of less than 1000, those being ISCED 3 (without distinction A-B-C, 2 years or more),

ISCED 3C for 2 years or more, ISCED 2, ISCED 5A that is a masters' degree, ISCED 1, and ISCED 4A-B. Only three educational requirements are below 100, that is ISCED 4 (without distinction A-B-C), ISCED 3C for shorter than 2 years, and ISCED 6. Jobs that require an education of below ISCED 1 or no formal education has a count of 417.

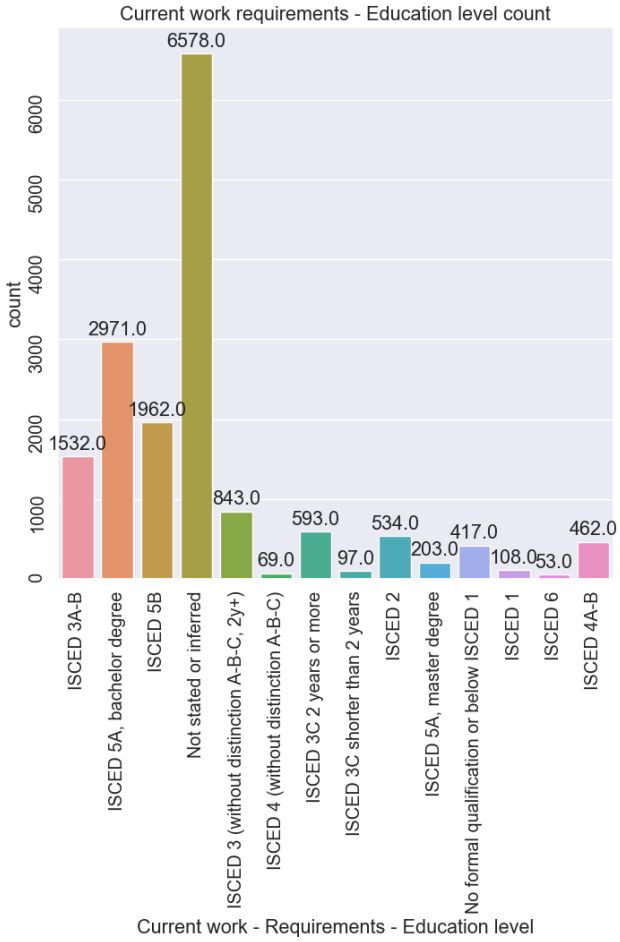

Fig. 13. Countplot of current work educational requirements

A mismatch in qualification can be seen particularly abundant in those that are over qualified for their job. Typically, this group consists majorly of weak performers with a count of 4139, while moderate performers are the second highest count with 3246. Meanwhile at risk performers have a count of 1767, while the category with the least count, strong performers, has a total of 615.

Those who are of equal status are those under the weak and moderate performers category, with a count of 1950 and 1866 respectively. At risk performers totals up to 704 while strong performers have a count of 461. For under qualified workers, weak performers total up to a count of 700, moderate performers have a count of 574, and at risk performers total up to 301. On the other hand, strong performers have the smallest count in this category, that is a count of 99.

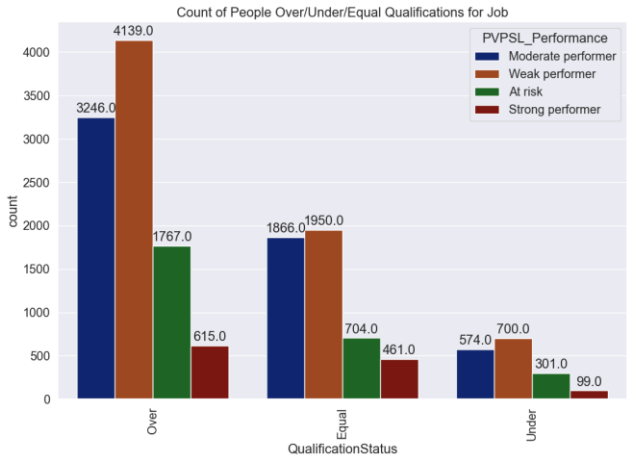

Fig. 14. Countplot people who are over/under/equal qualified for their job

E. Investigating the Data

The data was collected from the official PIAAC website. The question of 'why do Asians work so hard?' arises from the initial exploration of the dataset, revealing how the working hours in Asian countries continue to rise even as the workers age. This process is done in parallel to conducting the literature review and identifying the research gap. Once these are done, the data will be pre-processed. The next course of action includes exploratory data analysis, modelling the data, and visualisation of the results.

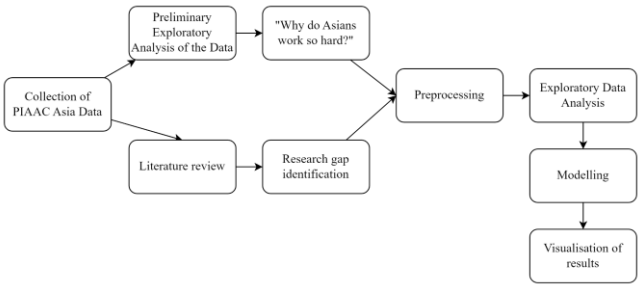

Fig. 15. FYP1 workflow

The data modelling workflow begins by separating the pre-processed data into several smaller datasets. Those who are dictated as working either over and under the average working hours will be separated and fed into three different machine learning models. The subsequent feature importances will then be extracted. Overqualified workers, under qualified workers, and equal qualified workers will be portioned out into their own datasets as well. The association rule mining will then be applied to each dataset.

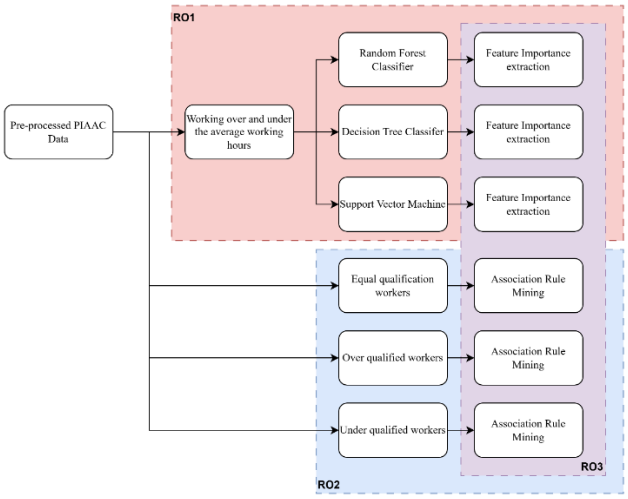

Fig. 16. Workflow of PIAAC data modelling

V. Chapter 6 Evaluation of Findings

A. *Research Objective 1*

The first research objective requires investigation of those working over the required working hours and under the required working hour. To recap, the objective is "investigating why people work more or fewer hours than required."

In looking at the working hours worked in each country, we observe that Japan, South Korea, Kazakhstan, and Singapore have 44%, 55%, 62%, and 43% of workers working over the average working hours. Meanwhile, only 24% and 21% of South Korean and Kazakhstan workers work under the average hours, while Japan and Singapore have 30% and 31% of workers working under the average hours. The remaining workers work within the average working hours.

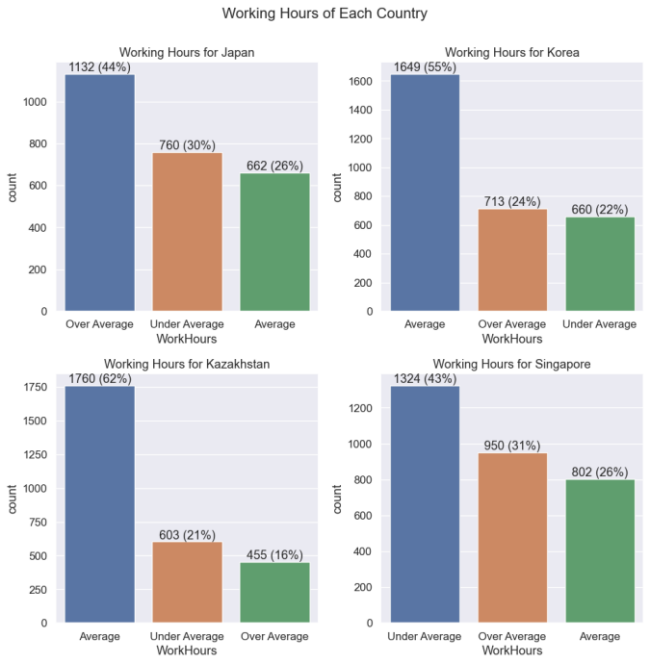

Fig. 17. Working Hours of Each Country

1) *Train-Test-Split*

Results from the train-test-split showcased that in terms of accuracy scores alone, the best parameters for the Random Forest Classifier were using the gini criterion with an n_val of 60 using a 15% test size to achieve a 0.757 accuracy. The decision tree classifier on the other hand was tied between using entropy and log loss criterion, both achieved the same score of 0.676 using a 5% test size. Meanwhile, the SVM showed best results when using the linear kernel with a 15% test size, achieving an accuracy of 0.730.

TABLE III. TRAIN-TEST-SPLIT RESULTS

Model	Criterion/Kernel	Test size	Accuracy (rounded up to 3 decimals)	N_val (for Random Forest Classifier)
Random Forest Classifier	Gini	0.15	0.757	60
	Entropy	0.15	0.753	70
	Log_loss	0.15	0.753	70
Decision Tree Classifier	Gini	0.05	0.667	-
	Entropy	0.05	0.676	-
	Log_loss	0.05	0.676	-
Support Vector Machine (SVM)	Linear	0.15	0.730	-
	Poly	0.2	0.626	-
	RBF	0.1	0.627	-
	Sigmoid	0.1	0.547	-

Delving deeper into the accuracy scores, classification matrices were applied to each of the models' best performers. The Random Forest Classifier with gini criterion showcases the best overall performance, scoring the highest precision, recall, and f1-score for both the 'Over Average' and 'Under Average' class labels with scores within the range of 0.73 and 0.78, with a support of 483 for 'Over Average' and 507 for 'Under Average'. It also achieved the highest AUC ROC score that is 0.828529. The second best model goes to SVM with linear kernel. Precision, recall, and f1-scores for both classes were within the range of 0.70 and 0.76, while the support scores were similar to the Random Forest Classifier. The AUC ROC score achieved was 0.804195. Both Decision Tree Classifiers present similar results, resulting in them both being the least performing model in the experiment. Scores for precision, recall, and f1-score were between the range of 0.64 and 0.70, while supports for both were below 156 and 174 for the 'Over Average' and 'Under Average' classes respectively. Both models achieved low scores for AUC ROC that is 0.676945.

TABLE IV. TRAIN-TEST-SPLIT EVALUATION

Model		Random Forest Classifier (Gini)	Decision Tree Classifier (Entropy)	Decision Tree Classifier (Log_loss)	SVM (Linear)
Over Average	**Precision**	0.736328	0.644970	0.644970	0.708494
	Recall	0.780538	0.698718	0.698718	0.759834
	F1-Score	0.757789	0.670769	0.670769	0.733267
	Support	483	156	156	483

Under Average	**Precision**	0.778243	0.708075	0.708075	0.754237
	Recall	0.733728	0.655172	0.655172	0.702170
	F1-Score	0.755330	0.680597	0.680597	0.727273
	Support	507	174	174	507
AUC ROC Score		0.828529	0.676945	0.676945	0.804195

Feature importances were then extracted from the models with the highest accuracy scorings. Our findings reveal that all the models agree that 'Current status/work history - Subjective status' takes the most priority in determining whether an individual is classed as working over or under the average working hours. The Random Forest Classifier, and both Decision Tree Classifiers agree up to the first 6 features, that is the gender, ICT skill at home and at work, the area of study for an individual's highest education, and the language spoken at home. In the case of the SVM linear kernel model, the findings show that being a native speaker, country of origin, the level of computer use at work, how often an individual conducts manual labour for longs at work, and the ISCOSKIL 4, are determining factors in the working hours of an individual.

TABLE V. TRAIN-TEST-SPLIT TOP FEATURE IMPORTANCE

Random Forest Classifier (Gini)		**Decision Tree Classifier (Entropy)**		**Decision Tree Classifier (Log_loss)**		**SVM (Linear)**	
Feature	**Importance**	**Feature**	**Importance**	**Feature**	**Importance**	**Feature**	**Importance**
Current status/work history - Subjective status	0.104349	Current status/work history - Subjective status	0.093441	Current status/work history - Subjective status	0.093441	Current status/work history - Subjective status	0.388185
GENDER_R	0.039295	GENDER_R	0.031380	GENDER_R	0.031380	NATIVESPEAKER	0.346316
ICTHOME	0.030366	ICTHOME	0.031114	ICTHOME	0.031114	CNTRYID	0.165858
ICTWORK	0.028993	ICTWORK	0.028774	ICTWORK	0.028774	Skill use work - ICT - Computer - Level of computer use	0.150191
Education - Highest qualification - Area of study	0.024812	Education - Highest qualification - Area of study	0.024597	Education - Highest qualification - Area of study	0.024597	Skill use work - How often - Working physically for long	0.113012
LNG_HOME	0.021547	LNG_HOME	0.021143	LNG_HOME	0.021143	ISCOSKIL4	0.108504

2) Cross-Validation Implementation

The best accuracy results from the cross validations showed that the best number of folds were 2, 3, 4, 7, 9, and 10. The Random Forest Classifier performed its best with an n_val of 100 for each criterion, with the gini criterion obtaining the highest accuracy compared to the remaining two that is an accuracy of 0.760. As for the Decision Tree Classifier, the gini criterion performed best, obtaining an accuracy of 0.678. Finally, the SVM model showcased that the linear kernel outperformed the other kernels wit an accuracy score of 0.742.

TABLE VI. CROSS VALIDATION ACCURACY RESULTS

Model	**Criterion/Kernel**	**Folds**	**Accuracy (rounded up to 3 decimals)**	**N_val (for Random Forest Classifier)**
Random Forest Classifier	Gini	7	0.760	100
	Entropy	4	0.759	100
	Log_loss	4	0.759	100
Decision Tree Classifier	Gini	3	0.678	-
	Entropy	3	0.672	-
	Log_loss	3	0.672	-
Support Vector Machine (SVM)	Linear	9	0.742	-
	Poly	9	0.627	-
	RBF	10	0.614	-
	Sigmoid	2	0.514	-

Classification matrices were then applied to the models with the highest accuracies. Here, we observe that both the Random Forest Classifier with gini criterion and the SVM with linear kernel were the best suited models for classifying individuals with over and under working hours. Each of their precision, recall, and f1-scores are above 0.7, while their AUC ROC score is above 0.8. Meanwhile, the decision tree classifier scored above 0.6 but below 0.7 in each category, denoting it as the lesser effective model. Each model however, scored 3250 for over average support, and 3347 for under average support.

TABLE VII. CROSS VALIDATION CLASSIFICATION RESULTS

Model		**Random Forest Classifier (Gini)**	**Decision Tree Classifier (Gini)**	**SVM (Linear)**
Over Average	**Precision**	0.741664	0.673331	0.721508
	Recall	0.787077	0.673538	0.777231
	F1-Score	0.763696	0.673435	0.748334
	Support	3250	3250	3250
Under Average	**Precision**	0.780178	0.682905	0.766150
	Recall	0.733791	0.682701	0.708694
	F1-Score	0.756274	0.682803	0.736303
	Support	3347	3347	3347
AUC ROC Score		0.839633	0.678120	0.815173

The features were then extracted from the models with the highest accuracies. Random Forest Classifier with gini criterion and Decision Tree Classifier with gini criterion both shared their top 3 features, those being the educational level requirements of an individual's current workplace, the gender of an individual, and how often teaching was involved in the workplace. Meanwhile, the SVM with linear kernel had a different set of features for the first top 3 attributes. Those

being whether the individual was a native speaker of the country they took the PIAAC assessment in, their current work status, and the country they took the assessment in.

TABLE VIII. CROSS VALIDATION TOP FEATURE IMPORTANCE

Random Forest Classifier (Gini)		**Decision Tree Classifier (Gini)**		**SVM (Linear)**	
Feature	**Importance**	**Feature**	**Importance**	**Feature**	**Importance**
Current work - Requirements - Education level	0.100609	Current work - Requirements - Education level	0.189814	NATIVESPEAKER	0.395082
GENDER_R	0.038942	GENDER_R	0.057833	Current status/work history - Subjective status	0.394827
Skill use work - How often - Teaching people	0.030630	Skill use work - How often - Teaching people	0.035822	CNTRYID	0.179803

B. Research Objective 2

To answer the second research objective, that is, "conducting association rule analysis to identify factors influencing the mismatch between highest academic qualification and job requirements.", requires diving into the rules generated by the association rule mining model. This involves a thorough analysis of the rules and observing the disparities between an individual's highest educational qualification and the qualifications required for their employment.

The rule analysis will cover consequents that have at least the qualification status, and rules where the consequents contain only the qualification status. The qualification status here being the type of disparity between the highest education and the qualification required by their employer.

1) Over qualified

The model obtained 403204 rules where the consequents contained an occurrence of the qualification status alongside other rules. Of those rules, 30 were found that showcased a dependant relationship between the antecedents and the consequent and has a support and confidence of over 0.9.

The rules found show a lift between the range of 1.0001865341172709 to 1.051924159103957. The range of confidence for those rules is from 0.9339764201500536 to 0.9829419583517944. Meanwhile, the support showcased a range of 0.901323955316508 to 0.917873396772859.

8 unique antecedents were found, and from those were 5 unique attributes. Those being a native speaker of the country where the induvial took the assessment, being born in the country where the individual took the assessment, having used a computer in their everyday life, having a count of 1 to 10 employees working under you, formal education being related to the job,

Meanwhile, 9 unique consequents were found alongside the qualification status. From these were 5 unique attributes, those are being born in the country where the assessment was taken, being a native speaker of the country where the assessment was taken, having used a computer in everyday life, having a count between 1 to 10 employees working for you, and having a formal education being for job related reasons.

For rules where there was only one consequent, 39419 rules were found. 25 rules were found that had a confidence and support of 0.9. Those rules had a lift and confidence of 1, while the support ranges from 0.901323955316508 to 1.

25 unique antecedents were found and from these were 7 unique attributes. These include the individual having used a computer in everyday life, having a count between 1 to 10 employees working for the individual, formal education qualification being related to the job, being a native speaker of the country where the assessment was taken, being born in the country where the assessment was taken, managing a count between 1 to 5 employees, and current work status being paid work be it a job or a business.

The 5 common antecedents attributes found between the two are formal education being job related, having used a computer in everyday life, being born in the country where the assessment was taken place, having a count between 1 to 10 employees working for you, and being a native speaker of the country where the assessment was taken.

2) Under qualified

For consequents that include attributes besides the qualification status, 638332 rules were generated. These rules were further refined to only include those with a lift over 1, a support and confidence of over 0.9. 54 rules fit these criteria, ranging a lift between 1.08224924657898 to 1.08224924657898. Meanwhile, the confidence had a range between 0.983745123537061 to 0.9967061923583662, while the support had a solid value of 0.9059880239520958.

16 unique antecedents were gathered. The unique attributes include being a native speaker in the country where the assessment was taken, having used a computer in everyday life, being born in the country where the assessment was taken, having a count of 1 to 10 employees working for you, and they are an employee at work.

Meanwhile, 16 unique consequents found. The 5 unique attributes for these are being born in the country where the assessment was taken place, being a native speaker of the country where the assessment was taken place, being an employee, having used a computer in everyday life, and having a count of 1 to 10 employees working for you.

For rules where the only consequent is the qualification status, 50167 rules were generated. Further refining the rules

to have a support of above 0.9 leaves us with 55 rules. These rules have a lift and confidence of 1, while the support ranges from 0.9005988023952096 to 1.

From these rules, 55 unique antecedents were extracted. The 7 unique attributes from these include being an employee, having used a computer in everyday life, having a count 1 to 10 employees working for you, having a formal educational qualification for job related reasons, current work status is paid work for a job or business, being a native speaker of the country where the assessment is taken place, and having experience with a computer in everyday life.

The common antecedents' unique attributes found were having used a computer in everyday life, being a native speaker of the country where the assessment was taken, being an employee, having a count of 1 to 10 employees working for you, and being born in the country where the assessment was taken.

3) Equal qualifications

For those with equal qualifications, a total of 797219 rules were generated. Further refining the rules to only include those with a confidence and support of over 0.9 and lift of over 1 gives us 108 rules. The lift from these rules ranges from 1.0023796923457946 to 1.0746374096900175. Meanwhile, the confidence for these ranges from 0.9503027771977448 to 0.9863466196872936 while the support ranges from 0.9019331453886428 to 0.9164317358034636.

There were 34 unique antecedents found, and the unique attributes for these include being a native speaker of the country where the assessment was taken, being an employee, being born in the country where the assessment was taken, having 1 to 10 employees working for you, had used a computer in everyday life, formal education qualification being for work related reasons, and current paid work being a job or a business.

4) Comparison of Association Rule Mining Results

An intersect, that is the common elements, of these rules reveal that 3 most common attributes for antecedents that have several consequent values are being born in the country where the assessment was taken place, having used a computer in everyday life, and being a native speaker of the country where the assessment took place.

Meanwhile, there are 6 unique attributes for antecedents where the qualification status was the only consequent. Those are being born in the country where the assessment was taken place, formal education being related to the job, having used a computer in everyday life, count of employees working for you is between 1 to 10, and being a native speaker of the country where the assessment is taken place.

The common attributes present in both these results are being born in the country where the assessment was taken, having used a computer in everyday life, and being a native speaker of the country where the assessment was taken place.

The intersects between these rules provide an idea of what commonly occurs between each of these classes, which may indicate that it is simply a common trait and does not influence how an individual addresses their skill gap.

Rules and attributes that are unique from other classes could reveal insight for what leads to being over qualified, under qualified, and equal qualified.

A look into the antecedents where there are other values in the consequents besides the qualification status reveals several pieces of information. For those who are over qualified for their current careers, the attribute found is that they have used a computer in everyday life and their current work involves having a count between 1 to 10 people working for the individual. Meanwhile, for those who are of equal qualification, the attributes found are current paid work being a paid job or business, formal education qualification being job related, having used a computer in everyday life, being an employee, and having a count between 1 to 10 people working for the individual. The under qualified rules were not able to obtain rules that were not present in the over qualified and equal qualifications rules.

As for the antecedents where the consequents is only the qualification status, several attributes were found for all 3 categories. The over qualified attributes include those who were born in the country where the assessment was taken, having a formal education related to their job, having used a computer in everyday life, managing a count between 1 to 5 other employees at work, and being a native speaker of the country where the assessment was done. Meanwhile, the under qualified attributes were having experience with a computer in everyday life, being born in the country where the assessment was taken, having paid work that is a job or a business, formal education being job related, having used a computer in everyday life, being an employee, having a count between 1 to 10 people working for the individual, and being a native speaker of the country where the assessment was taken. Finally, the equal qualifications attributes include having experience with a computer in everyday life, being born in the country where the assessment was taken, having paid work be it a job or a business, formal education qualification being related to the job, having used a computer in everyday life, being an employee, having a count of 1 to 10 people working for you, and being a native speaker of the country where the assessment was taken.

C. Research Objective 3

The third research objective involves investigating how individuals address the skill gap when they lack the necessary qualifications in their current profession. When looking at the results for factors attributed to working hours, we see that the common factors include gender. Other factors such as current work status, skill use at home and at work, the area of study of a participant's highest qualification, the language spoken at home, the education level required at work, and how often they teach people, are also attributes to consider. From this, we can discern that a person's working hours are affected by these attributes, and are possible

contributions towards how these individuals address the skill gap.

TABLE IX. WORKING HOURS FACTORS

	Unique attributes
Train-test-split	• Current status/work history - Subjective status • GENDER_R • ICTHOME • ICTWORK • Education - Highest qualification - Area of study • LNG_HOME
Cross validation	• Current work - Requirements - Education level • GENDER_R • Skill use work - How often - Teaching people

As for the association rules, those with the highest possible lift, confidence, and support value is analysed.

For those who are overqualified, there is a consistent wherein factors such as the local citizenship of the individual, whether they are a native speaker of that country and having used computer in their everyday life correlates to being over qualified.

Under qualified individuals are associated with the same factors as those who are over qualified. However, additional factors such as having a count of 1 to 10 employees and considering themselves an employee would lead to being under qualified. This indicates that individuals who are lacking skills to fulfil their occupational requirements are usually employees who are trying to manage a team of people to compensate. Unique attributes from feature importance and association rule mining.

TABLE X. TOP ASSOCIATION RULES

	Association Rules
Over qualified	Background - Born in country_Yes -> Qualification_Status_Over, NATIVESPEAKER_Yes Skill use everyday life - ICT - Ever used computer_Yes, Background - Born in country_Yes -> Qualification_Status_Over, NATIVESPEAKER_Yes Background - Born in country_Yes -> Qualification_Status_Over, NATIVESPEAKER_Yes, Skill use everyday life - ICT - Ever used computer_Yes NATIVESPEAKER_Yes -> Qualification_Status_Over, Background - Born in country_Yes Skill use everyday life - ICT - Ever used computer_Yes, NATIVESPEAKER_Yes -> Qualification_Status_Over, Background - Born in country_Yes NATIVESPEAKER_Yes -> Qualification_Status_Over, Background - Born in country_Yes, Skill use everyday life - ICT - Ever used computer_Yes
Under qualified	Background - Born in country_Yes -> Qualification_Status_Under, NATIVESPEAKER_Yes Background - Born in country_Yes -> Qualification_Status_Under, Skill use everyday life - ICT - Ever used computer_Yes, NATIVESPEAKER_Yes, Current work - Employee or self-employed_Employee Current work - Employee or self-employed_Employee, Background - Born in country_Yes -> Qualification_Status_Under, Skill use everyday life - ICT - Ever used computer_Yes, NATIVESPEAKER_Yes, Current work - Employees working for you - Count_1 to 10 people Background - Born in country_Yes, Current work - Employees working for you - Count_1 to 10 people -> Qualification_Status_Under, Skill use everyday life - ICT - Ever used computer_Yes, NATIVESPEAKER_Yes, Current work - Employee or self-employed_Employee Skill use everyday life - ICT - Ever used computer_Yes, Background - Born in country_Yes -> Qualification_Status_Under, Current work - Employee or self-employed_Employee, NATIVESPEAKER_Yes, Current work - Employees working for you - Count_1 to 10 people Current work - Employee or self-employed_Employee, Background - Born in country_Yes, Current work - Employees working for you - Count_1 to 10 people -> Qualification_Status_Under, Skill use everyday life - ICT - Ever used computer_Yes, NATIVESPEAKER_Yes Current work - Employee or self-employed_Employee, Skill use everyday life - ICT - Ever used computer_Yes, Background - Born in country_Yes -> Qualification_Status_Under, NATIVESPEAKER_Yes, Current work - Employees working for you - Count_1 to 10 people Skill use everyday life - ICT - Ever used computer_Yes, Background - Born in country_Yes, Current work - Employees working for you - Count_1 to 10 people -> Qualification_Status_Under, Current work - Employee or self-employed_Employee, NATIVESPEAKER_Yes Current work - Employee or self-employed_Employee, Skill use everyday life - ICT - Ever used computer_Yes, Background - Born in country_Yes, Current work - Employees working for you - Count_1 to 10 people -> Qualification_Status_Under, NATIVESPEAKER_Yes Background - Born in country_Yes -> Qualification_Status_Under, Skill use everyday life - ICT - Ever used computer_Yes, NATIVESPEAKER_Yes, Current work - Employees working for you - Count_1 to 10 people Background - Born in country_Yes, Current work - Employees working for you - Count_1 to 10 people -> Qualification_Status_Under, Skill use everyday life - ICT - Ever used computer_Yes, NATIVESPEAKER_Yes Skill use everyday life - ICT - Ever used computer_Yes, Background - Born in country_Yes -> Qualification_Status_Under, NATIVESPEAKER_Yes, Current work - Employees working for you - Count_1 to 10 people Skill use everyday life - ICT - Ever used computer_Yes, Background - Born in country_Yes, Current work - Employees working for you - Count_1 to 10 people -> Qualification_Status_Under, NATIVESPEAKER_Yes Current work - Employee or self-employed_Employee, Background - Born in country_Yes -> Qualification_Status_Under, Skill use everyday life - ICT - Ever used computer_Yes, NATIVESPEAKER_Yes Current work - Employee or self-employed_Employee, Background - Born in country_Yes -> Qualification_Status_Under, NATIVESPEAKER_Yes

	Skill use everyday life - ICT - Ever used computer_Yes, Background - Born in country_Yes -> Qualification_Status_Under, Current work - Employee or self-employed_Employee, NATIVESPEAKER_Yes Current work - Employee or self-employed_Employee, Skill use everyday life - ICT - Ever used computer_Yes, Background - Born in country_Yes -> Qualification_Status_Under, NATIVESPEAKER_Yes Background - Born in country_Yes -> Qualification_Status_Under, Current work - Employee or self-employed_Employee, NATIVESPEAKER_Yes, Current work - Employees working for you - Count_1 to 10 people Current work - Employee or self-employed_Employee, Background - Born in country_Yes -> Qualification_Status_Under, NATIVESPEAKER_Yes, Current work - Employees working for you - Count_1 to 10 people Background - Born in country_Yes, Current work - Employees working for you - Count_1 to 10 people -> Qualification_Status_Under, Current work - Employee or self-employed_Employee, NATIVESPEAKER_Yes Current work - Employee or self-employed_Employee, Background - Born in country_Yes, Current work - Employees working for you - Count_1 to 10 people -> Qualification_Status_Under, NATIVESPEAKER_Yes Background - Born in country_Yes -> Qualification_Status_Under, Skill use everyday life - ICT - Ever used computer_Yes, NATIVESPEAKER_Yes Skill use everyday life - ICT - Ever used computer_Yes, Background - Born in country_Yes -> Qualification_Status_Under, NATIVESPEAKER_Yes Background - Born in country_Yes -> Qualification_Status_Under, NATIVESPEAKER_Yes, Current work - Employees working for you - Count_1 to 10 people Background - Born in country_Yes, Current work - Employees working for you - Count_1 to 10 people -> Qualification_Status_Under, NATIVESPEAKER_Yes Background - Born in country_Yes -> Qualification_Status_Under, Current work - Employee or self-employed_Employee, NATIVESPEAKER_Yes Background - Born in country_Yes -> Skill use everyday life - ICT - Ever used computer_Yes, Current work - Employees working for you - Count_1 to 10 people, Current work - Employee or self-employed_Employee, Qualification_Status_Under, NATIVESPEAKER_Yes
Equal qualification	Background - Born in country_Yes -> Qualification_Status_Equal, NATIVESPEAKER_Yes Background - Born in country_Yes -> Qualification_Status_Equal, Skill use everyday life - ICT - Ever used computer_Yes, NATIVESPEAKER_Yes, Current work - Employee or self-employed_Employee Current work - Employee or self-employed_Employee, Background - Born in country_Yes -> Qualification_Status_Equal, Skill use everyday life - ICT - Ever used computer_Yes, NATIVESPEAKER_Yes, Current work - Employees working for you - Count_1 to 10 people Background - Born in country_Yes, Current work - Employees working for you - Count_1 to 10 people -> Qualification_Status_Equal, Skill use everyday life - ICT - Ever used computer_Yes,

	NATIVESPEAKER_Yes, Current work - Employee or self-employed_Employee Skill use everyday life - ICT - Ever used computer_Yes, Background - Born in country_Yes -> Qualification_Status_Equal, Current work - Employee or self-employed_Employee, NATIVESPEAKER_Yes, Current work - Employees working for you - Count_1 to 10 people Current work - Employee or self-employed_Employee, Background - Born in country_Yes, Current work - Employees working for you - Count_1 to 10 people -> Qualification_Status_Equal, Skill use everyday life - ICT - Ever used computer_Yes, NATIVESPEAKER_Yes Current work - Employee or self-employed_Employee, Skill use everyday life - ICT - Ever used computer_Yes, Background - Born in country_Yes -> Qualification_Status_Equal, NATIVESPEAKER_Yes, Current work - Employees working for you - Count_1 to 10 people Skill use everyday life - ICT - Ever used computer_Yes, Background - Born in country_Yes, Current work - Employees working for you - Count_1 to 10 people -> Qualification_Status_Equal, Current work - Employee or self-employed_Employee, NATIVESPEAKER_Yes Current work - Employee or self-employed_Employee, Skill use everyday life - ICT - Ever used computer_Yes, Background - Born in country_Yes, Current work - Employees working for you - Count_1 to 10 people -> Qualification_Status_Equal, NATIVESPEAKER_Yes Background - Born in country_Yes -> Qualification_Status_Equal, Skill use everyday life - ICT - Ever used computer_Yes, NATIVESPEAKER_Yes, Current work - Employees working for you - Count_1 to 10 people Background - Born in country_Yes, Current work - Employees working for you - Count_1 to 10 people -> Qualification_Status_Equal, Skill use everyday life - ICT - Ever used computer_Yes, NATIVESPEAKER_Yes Skill use everyday life - ICT - Ever used computer_Yes, Background - Born in country_Yes -> Qualification_Status_Equal, NATIVESPEAKER_Yes, Current work - Employees working for you - Count_1 to 10 people Skill use everyday life - ICT - Ever used computer_Yes, Background - Born in country_Yes, Current work - Employees working for you - Count_1 to 10 people -> Qualification_Status_Equal, NATIVESPEAKER_Yes Current work - Employee or self-employed_Employee, Background - Born in country_Yes -> Qualification_Status_Equal, Skill use everyday life - ICT - Ever used computer_Yes, NATIVESPEAKER_Yes Current work - Employee or self-employed_Employee, Background - Born in country_Yes -> Qualification_Status_Equal, NATIVESPEAKER_Yes Skill use everyday life - ICT - Ever used computer_Yes, Background - Born in country_Yes -> Qualification_Status_Equal, Current work - Employee or self-employed_Employee, NATIVESPEAKER_Yes Current work - Employee or self-employed_Employee, Skill use everyday life - ICT - Ever used computer_Yes, Background - Born in country_Yes -> Qualification_Status_Equal, NATIVESPEAKER_Yes Background - Born in country_Yes -> Qualification_Status_Equal, Current work -

	Employee or self-employed_Employee, NATIVESPEAKER_Yes, Current work - Employees working for you - Count_1 to 10 people Current work - Employee or self-employed_Employee, Background - Born in country_Yes -> Qualification_Status_Equal, NATIVESPEAKER_Yes, Current work - Employees working for you - Count_1 to 10 people Background - Born in country_Yes, Current work - Employees working for you - Count_1 to 10 people -> Qualification_Status_Equal, Current work - Employee or self-employed_Employee, NATIVESPEAKER_Yes Current work - Employee or self-employed_Employee, Background - Born in country_Yes, Current work - Employees working for you - Count_1 to 10 people -> Qualification_Status_Equal, NATIVESPEAKER_Yes Background - Born in country_Yes -> Qualification_Status_Equal, Skill use everyday life - ICT - Ever used computer_Yes, NATIVESPEAKER_Yes Skill use everyday life - ICT - Ever used computer_Yes, Background - Born in country_Yes -> Qualification_Status_Equal, NATIVESPEAKER_Yes Background - Born in country_Yes -> Qualification_Status_Equal, NATIVESPEAKER_Yes, Current work - Employees working for you - Count_1 to 10 people Background - Born in country_Yes, Current work - Employees working for you - Count_1 to 10 people -> Qualification_Status_Equal, NATIVESPEAKER_Yes Background - Born in country_Yes -> Qualification_Status_Equal, Current work - Employee or self-employed_Employee, NATIVESPEAKER_Yes Background - Born in country_Yes -> Qualification_Status_Equal, Skill use everyday life - ICT - Ever used computer_Yes, Current work - Employees working for you - Count_1 to 10 people, Current work - Employee or self-employed_Employee, NATIVESPEAKER_Yes

In observing all the rules gathered, the attributes were taken based on the unique attributes found in both the antecedents with multiple consequent values and a singular consequent value, instead of the unique antecedents. The common attributes found in all three classes is removed, leaving only those unique to the class

With these parameters, the association rule mining results showcase that for the over qualified workers, the common attributes are those born in the country of the where the assessment was taken, those having a formal education related to their job, having used a computer in their everyday life, having a count between 1 to 10 employees working for the individual, and being a native speaker of the country where the assessment was taken place. These attributes entail an over qualified individual is most often a local citizen, manage a small team at work, use a computer in their daily lives, and took an education for their current career.

Meanwhile, attributes that are associated with those who are under qualified include those born in the country where the assessment was taken place, having used a computer in their everyday life, having a count between 1 to 10 employees working for the individual, is an employee themselves, and is a native speaker of the country where the assessment was taken place. The attributes lead to a local who manages a team under their superior, and most likely uses a computer at home.

As for those with equal qualifications, we gather the attributes being born in the country where the assessment was taken place, being a native speaker of the country where the assessment was taken place, having a formal education due to job related reasons, receiving paid work be it a job or a business, having used a computer in their everyday life, having a count between 1 to 10 employees working under the individual, and being an employee. The description of a local citizen with a qualification suited for their career comes to mind. This individual is also an employee managing a small team, while working a paid job that may or may not be family business.

TABLE XI. UNIQUE ATTRIBUTES FROM TOP ASSOCIATION RULE MINING

	Unique attributes
Over qualified	Background - Born in country_Yes Education - Formal qualification - Reason job related_Yes Skill use everyday life - ICT - Ever used computer_Yes Current work - Employees working for you - Count_1 to 10 people NATIVESPEAKER_Yes
Under qualified	Background - Born in country_Yes Skill use everyday life - ICT - Ever used computer_Yes Current work - Employees working for you - Count_1 to 10 people Current work - Employee or self-employed_Employee NATIVESPEAKER_Yes
Equal qualification	Background - Born in country_Yes Education - Formal qualification - Reason job related_Yes Current status/work history - Current - Paid job or family business (DERIVED BY CAPI)_Yes, paid work one job or business Skill use everyday life - ICT - Ever used computer_Yes Current work - Employees working for you - Count_1 to 10 people Current work - Employee or self-employed_Employee NATIVESPEAKER_Yes

VI. CHAPTER 7 CONCLUSION

The first research objective, to examine factors that influence working over or under the required working hours, was conducted with the implementation of 6 machine learning models. Of these models, the Random Forest Classifier with gini criterion showed itself as the best fit model thus far. The common features extracted from all the models used indicated that gender was a primary factor in determining working hours. Meanwhile, other factors such as an individuals' ICT skills at work and at home, the area of study for their highest educational qualification, language spoken at home, their occupation's educational requirement level, and how often they teach people at work, are additional factors to consider.

The second objective was achieved by conducting an association rule analysis to identify factors influencing the mismatch between highest qualification and employment qualification requirement. The output highlights that those

who are not under qualified are associated with having an educational qualification that is related to their career. Meanwhile, equal qualification workers showcase that their current paid work is either a job or a business. Both these indicate that those who are under qualified tend to lack the educational requirements for their career. They also may not be recipients of paid work.

The final objective, that is to investigate how individuals address the skill gap when they lack the necessary qualifications in their current profession, was done by comparing the outputs obtained from the experiments conducted for this project. What is concluded is that in addressing the issue of how individuals address the skill gap when they lack the necessary qualifications in their current profession, the findings show that most often the difference between those who are under qualified and those who have equal qualifications are working a paid work be it a job or a business. On the other hand, the difference between those who are over qualified and those who are under qualified is that over qualified individuals tend to take a leadership role as opposed to a passive one.

To summarize, the project reveals that gender is a primary factor influencing individuals' working hours. The interplay between highest qualification, educational attainment, and professional role at work is crucial in addressing career skill gaps. It highlights the importance of aligning educational qualifications with career requirements and understanding the behaviours of those who exceed these skill gaps. These findings can aid in the development of policies and support mechanisms to enhance individuals' professional careers.

A. Limitations

Since the majority of researchers utilised software programs like STATA, the financial need for utilising these capabilities was not able to be fulfilled. Therefore, additional investigation was necessary to discover other approaches that were open source or freely available for use.

The size and complexity of the dataset proved itself robust as the association rule mining parameters were limited to ensure it would run smoothly. This is due to the hardware equipment used being unable to support the amount of memory required to run the model.

B. Future Plans

Additional improvements, specifically, the concept of categorising workers according to the ISCOSKIL4 classification, would further refine the scope of the project and help gain more insights into each of the specific classification details. This will provide a comprehensive study to compare and identify the specific categories of workers that are particularly prone to experiencing skill mismatch and exceeding the normal working hours. Here, the distinct demands and wants of various worker groups can be identified to determine the most effective approach to reducing the skill gap they possess.

REFERENCES

[1] D. Liao, Q. He and H. Jiao, "Mapping background variables with sequential patterns in problem-solving environments: An investigation of United States adults' employment status in PIAAC," Frontiers in Psychology, vol. 10, 27 March 2019.

[2] M. Pivovarova and J. M. Powers, "Do immigrants experience labor market mismatch? New evidence from the US PIAAC," Large-Scale Assessments in Education, vol. 10, no. 1, 2022.

[3] R. Hämäläinen, B. D. Wever, K. Nissinen and S. Cincinnato, "Understanding adults' strong problem-solving skills based on PIAAC," Journal of Workplace Learning, vol. 29, no. 7/8, pp. 537-553, 11 September 2017.

[4] D. S. Olsen and T. Tikkanen, "The developing field of workplace learning and the contribution of PIAAC," International Journal of Lifelong Education, vol. 37, no. 5, p. 546–559, 2018.

[5] A. Grotlüschen, C. Stammer and T. J. Sork, "People who teach regularly: What do we know from PIAAC about their professionalization?," Journal of Adult and Continuing Education, vol. 26, no. 1, 2020.

[6] Q. He, Q. Shi and E. L. Tighe, "Predicting Problem-Solving Proficiency with Multiclass Hierarchical Classification Using Process Data: A Machine Learning Approach," Psychological Test and Assessment Modeling, vol. 65, no. 1, pp. 145-178, 2023.

[7] C. Hahnel, U. Kroehne and F. Goldhammer, "Rule-based Process Indicators of Information Processing Explain Performance Differences in PIAAC Web Search Tasks," Large-scale Assessments in Education, vol. 11, no. 16, 19 May 2023.

[8] H. Komatsu and J. Rappleye, "Refuting the OECD-World Bank development narrative: was East Asia's 'Economic Miracle' primarily driven by education quality and cognitive skills?," Globalisation, Societies and Education, vol. 17, no. 2, pp. 101-116, 2019.

[9] J.-W. Lee, D. W. Kwak and E. Song, "Can older workers stay productive? The role of ICT skills and training," Journal of Asian Economics, vol. 79, April 2022.

[10] D. H. Lim, H. Ryu and B. Jin, "A latent class analysis of older workers' skill proficiency and skill utilization in South Korea," Asia Pacific Education Review, vol. 21, p. 365–378, 28 May 2020.

[11] J. Yoon, E. J. Hur and M. Kim, "An analysis of the factors on the problem-solving competencies of engineering employees in Korea," Sustainability, vol. 12, no. 4, 2020.

[12] C.-Y. Jyung, Y. Lee, S. Park, E. Cho and R. Choi, "Factors Affecting Employees' Problem-Solving Skills in Technology-Rich Environments in Japan and Korea," Sustainability, vol. 12, no. 17, p. 7079, 2020.

[13] U. Tolganay, Effects of Skills Mismatches on Job Satisfaction in Kazakhstan: Evidence from PIAAC Data, 2021.

[14] J.-W. Lee and D. Wie, "Returns to education and skills in the labor market: Evidence from japan and korea. Asian Economic Policy Review," Asian Economic Policy Review, vol. 12, no. 1, p. 139–160., 2017.

[15] OECD, "Technical Report of the Survey of Adult Skills (PIAAC) (3rd Edition)," 2019.

[16] A. Garg, "Complete guide to Association Rules (1/2)," 7 February 2019. [Online]. Available: https://towardsdatascience.com/association-rules-2-aa9a77241654 .

[17] J.-W. Lee, J.-S. Han and E. Song, "The effects and challenges of vocational training in Korea," International Journal of Training Research, vol. 17, no. 1, p. 96–111, 2019.