I. Executive Summary

Due to the impact of the COVID-19 pandemic, remote work has become the preferred working condition. Many employers have found it challenging to transition back to in-person work. Remote work allows the benefits of diversity, inclusion, and financial savings. However, this may be at the cost of reinforcing gender norms and potential mental burnout of employees, especially for younger employees. Considering these factors, this paper focuses on determining 1) *What factors affect employees' perceived productivity in remote working conditions?* and 2) *How do employers benefit from remote work systems?*

The data originated from two surveys conducted by the authorities of New South Wales to understand remote workers' experiences during the COVID-19 pandemic. The first survey was held in August and September 2020, and the second in 2021. The datasets were combined based on similar questions, resulting in 29 columns. Perceived productivity was the output variable and was categorised into three class variables: low, same, and high. The input variables were a mix of numeric (i.e., percentage of remote work in 2020 and preferred remote work percentage in 2020) and categorical variables (i.e., gender and occupation description). The data was split into training (70%) and validation (30%) subsets. Any rows with missing values, columns deemed unnecessary, and individuals 67 years old or older were removed. Categorical variables were encoded to numbers, and any scaled responses were transformed into ordinal scales. Outliers were identified using the 3-standard-deviation rule and replaced with column medians. The final data, after transformation, contained 2,804 rows and 29 columns.

This study used random forests, feature importance analysis, and partial dependence plots to determine what factors influence employees' productivity when working remotely. Random forests were used to classify perceived productivity into the three output classes. When testing, the random forest obtained an accuracy of 60,69% on the validation set, 77,42% on the training set, a mean absolute error with a score of 0,52, and an ordinal accuracy of 83,39%. Attempts were made to solve the issue of the class imbalance seen within the random forest model. However, none of the other models or sampling techniques created better results. Feature importance was used to discover the top ten factors influencing productivity in remote work settings. Finally, partial dependence was conducted to better understand the connection certain factors had on productivity; this was done by isolating the relationship between one factor and productivity.

The final results for the top factors that influence productivity in a remote work setting in descending order are the following: *employees with a strong preference for remote work, age of employees, the number of hours a day employees spend remote working, in-person commuting times, employees who found collaboration easier with remote work, employees who had experience with remote work during 2020, how many hours an employee can use with family or on domestic responsibilities.* Deeper analysis found that an individual's occupation (i.e., Managerial or Technician) influenced their attitude towards remote work.

Overall, productivity rose with remote work for Managerial Positions, Corporate and Public Sector Professionals, and Sales Workers. It is suggested that for these occupations to improve productivity with virtual work further, employers should *promote flexible work agreements, invest in tools that allow easier online collaboration, encourage work-life balance, aid employees in what barriers make remote work difficult, and provide tailored support for different age groups.*

II. Introduction

The aftermath of the COVID-19 pandemic completely changed work culture. Remote work has become the preferred method to get people involved in the workspace. After a few years of being either fully remote or hybrid, many employers have found it rather impossible to transition back into in-person work.

Remote work enables employees in distant locations or with caretaking responsibilities to remain accessible. Some studies find that women, particularly those balancing caregiving responsibilities, favour remote work options for greater flexibility (Tenakwah and Watson, 2024, p. 135). Remote work also allows employers to hire from more productive talent pools without the location concern (Tenakwah and Watson, 2024, pp. 137-8). However, there is a concern that remote work still upholds gender inequality, as it is found that women tend to earn less than their male counterparts and experience job insecurity. Women also might experience burnout due to the struggle of balancing fulltime work with caretaking activities (Höcker et al., 2024, p. 328).

As more employees work from home, a business could save between $8,500 to $13,000 per employee per year because of the reduced need for office space (Beck and Hensher, 2022, pp. 280). Employees also benefit either by saving time and money from no longer needing to commute or by no longer needing to live in or near a high rent city where an office would have been located (Beck and Hensher, 2022, pp. 280). However, a drawback to remote working is that employers find that centralised office setups are essential for fostering creativity, innovation, and employee engagement (Tenakwah and Watson, 2024, pp. 137). Also, younger employees are more likely to experience burnout because they "are more likely to suffer from feelings of isolation, anxiety, depression and stress as a result of remote work." These younger employees also might not have a suitable workspace for remote work (Höcker et al., 2024, pp. 328). This can lead to the blurring of work-life balance, which can also result in burnout.

Considering these factors, this paper focuses on determining 1) *What factors affect employees' perceived productivity in remote working conditions?* and 2) *How do employers benefit from remote work systems?* Explanatory modelling was conducted to answer these problems.

III. Data Processing

The dataset originates from two surveys conducted by the authorities of New South Wales (NSW), in Australia, to understand remote workers' experiences during the COVID-19 pandemic. The first survey, conducted in August and September 2020, captured insights during the height of the lockdown, while the second survey, conducted in 2021, provided a follow-up on evolving remote work patterns. The surveys collectively captured responses from 3000 participants, focusing on various aspects of remote work, such as organisational readiness, worker preferences, productivity, and time allocation between professional and personal responsibilities. Due to differing survey designs, the datasets were combined based on similar questions in two surveys, resulting in a final dataset with 29 columns after harmonising and standardising the variables.

Input Variables:

- **Numeric**: Examples include rw_percentage_2020 (percentage of remote work in 2020), preferred_rw_2020 (preferred remote work percentage in 2020), preferred_rw_percentage_future (future preference for remote work), and age.
- **Categorical**: Examples include gender, industry_desc (industry description), occupation_desc (occupation description), and household composition.

Output Variable: Remote vs. In-person Productivity - a 3 class measure of the productivity difference between remote and in-person work. The initial data had responses ranging from -100% to 100%, in increments of 10%. Respondents were classified as "more productive" for positive values, "same productivity" for a value of zero and "less productive" for negative values. **Data Preparation**:

The dataset was split into **training (70%)** and **validation (30%)** subsets for analysis. Preprocessing was done to ensure data consistency, handle missing values, and prepare the dataset for analysis. Key steps included: **1) Data Cleaning**: Rows with missing values were removed to maintain data integrity. Irrelevant columns were dropped, and the remaining columns were renamed for clarity. **2) Category Merging and Standardisation**: Categorical variables, such as industry descriptions, were combined by merging similar categories. Likert-scale responses and other ordinal variables were mapped to ordinal scales for analysis. **3) Derived Features**: The age variable was derived from respondents' birth years, and entries with extreme age values (67 and older) were excluded, as 67 is the retirement age in Australia (Department of Social Services, 2024). **4) Outlier Management**: Outliers in numerical columns were identified using the 3-standard-deviation rule and replaced with column medians. **5) Final Dataset**: After preprocessing, the dataset contained 2,804 rows and 29 columns. A screenshot of the final dataset has been added to the Appendix under Figure 1 for better visualisation.

IV. Methodology

This project builds on a project conducted by the Authorities of New South Wales to assess the economic impact of remote work for the region, why remote work is here to stay and inform employers about its benefits and how to adapt. However, this study only relied on descriptive and diagnostic analytics and did not try to identify the profile of employees who perceive productivity increases due to remote working. This is the gap that this paper is trying to address with the help of predictive analytics. The aim of this work is to build a model to predict if an individual, based on their personal factors, is going to be more productive, same productivity or less productive working remotely.

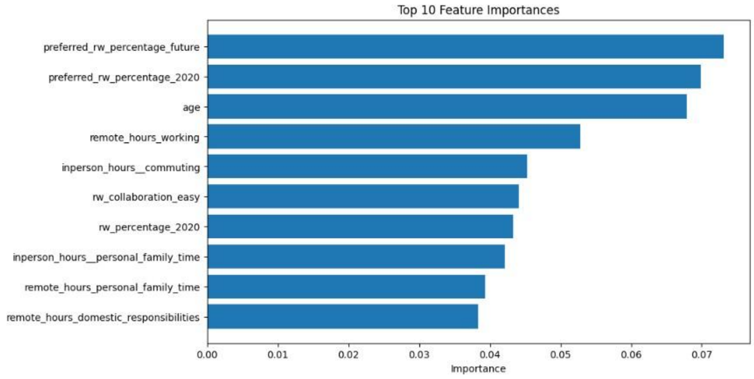
The classification method that seemed evident for this task was a Random Forest due to its properties and generally good performances. The model was trained on 70% of the data and was given a range of values to select from for its hyperparameters using five-fold cross validation[1]. Those are a set of parameters that can be adjusted to manage machine learning model training and are particularly efficient at preventing overfitting (Amazon, n.d.). The values were found using a random search with 30 iterations. This design choice was made due to the ability of random search to find equivalently good parameters and requiring less computational power (Bergstra & Bengio, 2012).

The Random Forest obtained an accuracy of 60,69% on the validation set compared to 77,42% on the training set. This represents an average performance. The Mean Absolutes Error is also computed with a score of 0,52 suggesting that on average our model is wrong by half a class. Furthermore, a value of 85,39% is obtained for the ordinal accuracy. This metric is usually used in cases when the number of levels in the ordinal data is larger (n>3), as it treats as correct all errors within a range of one class of the actual class. But in this context, it measures the percentage of "big" misclassification errors made by the model. Finally, to better visually represent those results the ROC is plotted[2]. This metric is usually used for two-class classification problems. Therefore, to make fit in this context three distinct curves were made, essentially transforming this into a two-class classification problem as treating each class against all the rest. The lowest Area Under the Curve (AUC) is 0,65 for the "same productivity" class and 0,7 on average, suggesting an acceptable level of differentiation.

[1] The list of the Hyperparameters and suggested values are provided in Table 1 of the Appendix. [2] The ROC Curve can be found in the Appendix under Figure 2.

Based on the Classification report[2] it can clearly be seen that the model suffers from class imbalance, as suggested by the low values for recall of the minority classes. Attempts were made to remediate this problem using other models such as Gradient Boost and XGBoost. Additionally, oversampling through algorithms that create synthetic data such as SMOTE were explored as well as under sampling. However, none of these approaches resulted in a better performance, this is partly explained by the ability of Random Forest to deal with class imbalance through its bagging mechanisms.

Nevertheless, the focus on this study is mostly to identify the profile of people who feel more productive using remote work. This is addressed in the next step by identifying the most important features for the model using "Feature Importance". What it essentially does is look at which features when used for the splits by the Random Forest resulted in a significantly better ability to discern the different classes using Gini's impurity coefficient (Stringer, 2018).



The final step is to plot the partial dependence plots for the "Top Features". This is important to understand the interactions at play, as "Top Features" does not tell anything about the connection between the predictor and the prediction. Partial dependence plots solve this problem by displaying the relationship of a feature with the predicted outcome. In this particular case as it is a classification problem the "y-axis" represents the probability across the feature values (x-axis) of belonging to a certain class. In case there are multiple classes, as in this context (3) the partial dependence plots must be computed for each target class (Molnar, 2024). Finally, the partial dependence plots for each "Occupation" class are plotted in order to spot potential patterns emerging from specific sectors of activity. Most of the insights presented in the report are derived from those plots. Full charts and additional analysis details are included in the Appendix.

V. Results

The focus of this report's insights lies on identifying the key drivers of productivity for Class 2 - 'More Productive.' Key takeaways derived from the Partial Dependence Plots (PDPs), illustrate that the relationships between key predictors and the likelihood of being classified as "More Productive" are detailed as follows: ***1) Employees who strongly prefer remote work*** *(both in 2020 and in the future) are significantly more likely to feel productive remotely.* ***2) Productivity rises with remote working hours****, peaking between 8 and 10 hours per day. Beyond this range, the effect plateaus, indicating a point of diminishing returns or issues within the reported data.* ***3) Employees with longer commutes are likely to report higher productivity during remote work****.* ***4) Collaboration ease*** *plays a pivotal role, and employees who find it easy to collaborate remotely are much more likely to report being more productive.* ***5) Productivity peaks for employees aged 30-45*** *and gradually declines beyond this range. This could reflect a balance of experience and adaptability, with younger employees possibly less experienced and older employees facing challenges adapting to remote work environments.* ***6) Productivity shows a nonlinear relationship with time spent on family and personal tasks.*** *Moderate time allocations enhance productivity, but excessive time (e.g., over 6-8 hours) decreases, likely due to distractions or challenges in balancing work and personal responsibilities.*

[2] The full classification report can be found under table 2 in the Appendix.

Furthermore, a comparison across different classes reveals a few distinct patterns: 1) Employees in Class 0 - 'Less Productive' exhibit a declining relationship with features such as remote work preference and collaboration ease. They tend to struggle more with remote work setups, possibly due to personal preferences or job-specific constraints. 2) Employees in Class 2 - 'More Productive' consistently demonstrate strong positive relationships with features such as remote work preference, collaboration ease, and reduced commuting time, emphasising the benefits of remote work for certain employee profiles.

More profound insights about remote work were analysed by grouping respondents across various occupation groups. This allows us to offer key insights tailored to each job role and account for the varying patterns and differences across occupations. These insights are outlined below:

1.Managerial and Professional Roles

These roles often involve independent and knowledge-based tasks which allows for greater flexibility for the inclusion and productivity of remote work. Longer commuting hours generally had a greater impact on decreasing overall productivity. Moderate time allocated to family and personal responsibilities (up to 4-6 hours/day), increases productivity but declines sharply beyond this.
Productivity peaks in the 35-45 age range, reflecting a balance of experience and adaptability.

2.Technicians, Trade and Sales Workers

The roles are generally more hands-on or customer-facing which affects these careers' overall productivity. Due to the necessity of on-site work, commuting may be unavoidable. Productivity declines after time allocated to family and personal responsibilities goes beyond 2-4 hours/day. Age exhibits similar trends with productivity as described above, however, to a lesser degree, productivity peaks between ages of 35-45, which may reflect the importance of other factors, such as work environment and tasks, as limiting factors in perceived productivity over one's age.

VI. Recommendations

Based on the findings, we conclude that rises in productivity due to remote work are primarily concentrated across three key occupation areas. Managerial Positions, Corporate and Public Sector Professionals. The following recommendations offer a few suggestions to further enhance remote work productivity specifically for these occupations as:

1. **Promote flexible work agreements** that allow employees to customise their work schedules based on individual preferences.
2. **Invest in tools that enhance the ease of remote collaboration**, as they are critical to supporting remote work productivity.
3. Encourage employees to **maintain a balance between work and personal responsibilities**. Moderate remote hours (e.g., 6-8 hours/day) and reasonable family/personal time allocations can maximise productivity.
4. For employees classified as less productive, **identify specific barriers** (e.g., job-specific constraints, lack of remote work infrastructure) **and implement targeted interventions**.
5. **Provide tailored support for different age groups**. For example, older employees may benefit from additional training in remote work technology, while younger employees may need guidance in managing work-life balance.

References

Academic

Alsulami, A., Mabrouk, F., & Bousrih, J. (2022). Flexible working arrangements and social sustainability: Study on Women Academics Post-covid 19. Sustainability, 15(1), 544. https://doi.org/10.3390/su15010544

Beck, M. J., & Hensher, D. A. (2022). Working from home in Australia in 2020: Positives, negatives and the potential for future benefits to transport and Society. Transportation Research Part A: Policy and Practice, 158, 271–284. https://doi.org/10.1016/j.tra.2022.03.016

Bergstra, J. and Bengio, Y. (2012) 'Random search for hyper-parameter optimization', *Journal of Machine Learning Research*, 13(2).

Höcker, M. C., Bachtal, Y., Voll, K., & Pfnür, A. (2024). Healthy, healthier, hybrid work: The Burnout-reducing potential of remote work and the mediating effect of work autonomy. International Journal of Workplace Health Management, 17(4), 319–334. https://doi.org/10.1108/ijwhm-02-2024-0036

Tenakwah, E. S., & Watson, C. (2024). Are we working from home or office? insights from Australia. Strategic HR Review, 23(4), 134–140. https://doi.org/10.1108/shr-03-2024-0017

Non-Academic

Amazon. (n.d.). *What is Hyperparameter Tuning? - Hyperparameter Tuning Methods Explained - AWS*. Amazon Web Services, Inc. https://aws.amazon.com/what-is/hyperparametertuning/#:~:text=computationally%20intensive%20process.-,What%20are%20hyperparameters%3F,set%20before%20training%20a%20model.

Department of Social Services. (2024). *Older Australians*. [online] Available at: https://www.dss.gov.au/older-australians.

Molnar, C. (2024, July 31). *8.1 Partial Dependence Plot (PDP) | Interpretable Machine Learning*. https://christophm.github.io/interpretable-ml-book/pdp.html

Stringer, S. (2018, July 28). Feature importance — what's in a name? - bigdatarepublic - Medium. *Medium*. https://medium.com/bigdatarepublic/feature-importance-whats-in-a-name79532e59eea3

Table 1 - *Suggested values for Hyperparameters*

Hyperparameter	Suggested Values
Number of estimators/trees	[<u>100</u>; 200; 300; 400; 500]
Maximum depth	[None; 10; 20; 30; <u>40</u>]
Minimum Samples Per Leaf	[<u>4</u>; 5; 6; 7; 8; 9; 10]
Maximum Number of Features Considered	[‘sqrt’; ‘<u>log2</u>’]

The best parameters found are underlined.

Table 2 - *Classification Report*

	Precision	Recall	F1-score	Support
Less productive (0)	0,62	0,14	0,23	148
Same productivity (1)	0,45	0,06	0,11	205
More productive (2)	0,61	0,98	0,75	489
Accuracy			0,61	842
Macro avg	0,56	0,39	0,36	842
Weighted avg	0,57	0,61	0,50	842

Figure 1 – *Snapshot of the Refined Dataset*

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	gender	dustry_de	upation_d	anization_	nage_oth	household	ears_at_jo	ro_or_regi	ercentage	encourage	repared_f	common_	ission_is_a	llaboratior	rw_percer
2	Female	Manufactı	Clerical ar	2	0	Couple wi	2	1	90	4	3	3	4	3	80
3	Male	Wholesale	Managers	0	1	Couple wi	2	0	20	3	3	3	3	3	20
4	Male	Electricity	Managers	3	1	One pare	2	0	50	2	3	3	2	3	60
5	Female	Professior	Professior	2	0	Couple wi	1	0	100	4	1	3	4	4	100
6	Male	Transport,	Managers	1	1	Couple wi	2	0	90	1	3	4	1	3	100
7	Male	Retail Trac	Sales Wor	2	0	Single per	1	0	70	4	4	4	4	4	50
8	Male	Financial ¿	Clerical ar	3	0	Couple wi	2	0	100	4	3	4	0	4	100
9	Male	Manufactı	Managers	1	1	Couple wi	0	0	90	4	4	3	3	4	80
10	Male	Administr	Clerical ar	3	0	Single per	2	0	60	4	3	4	3	3	70

	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
1	w_percen	mployer_	employer	uld_have	ty_remote	hours_	_co	n_hours_	s_person	domestic	hours_cor	e_hours_w	s_persona	domestic_	age
2	90	1	1	0	2	2	8	2	2	0,5	8	3,5	2	49
3	20	3	3	3	0	2	7	3	3	0	7	3	3	49
4	60	3	3	3	0	6	8	6	5	5	2	7	7	39
5	100	3	4	3	1	1	9	1	2	0	9	3	2	34
6	60	0	0	0	2	1	8	3,5	2	0	6	4	3	30
7	50	3	3	3	1	1	8	2	2	0	7	3	3	32
8	90	3	4	4	2	2	8	4	1	0	8	5	3	48
9	80	3	3	4	0	1	8	4	3	0,4	6	4	6	29
10	20	2	2	1	2	0	8	3,5	4	0	8	4	1	44

This screenshot is a snapshot of the data after being processed. The raw and refine datasets were both made available in the GitHub repository.

Figure 2 – *Roc Curve*

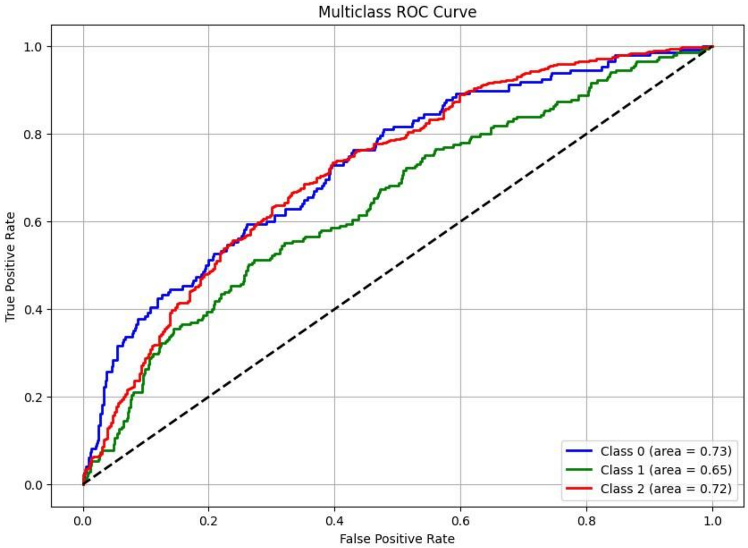


Figure 3 – *Feature Importance*

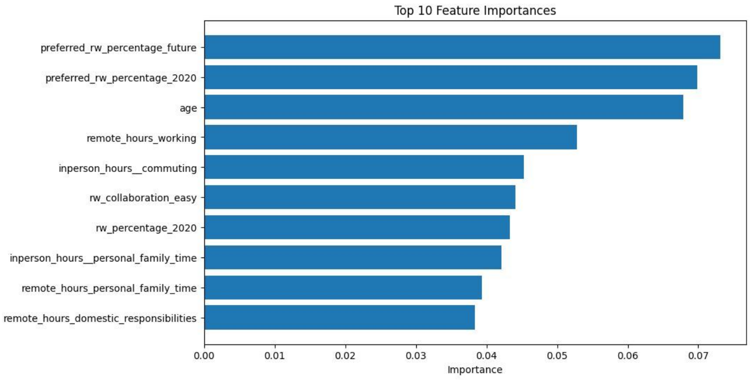
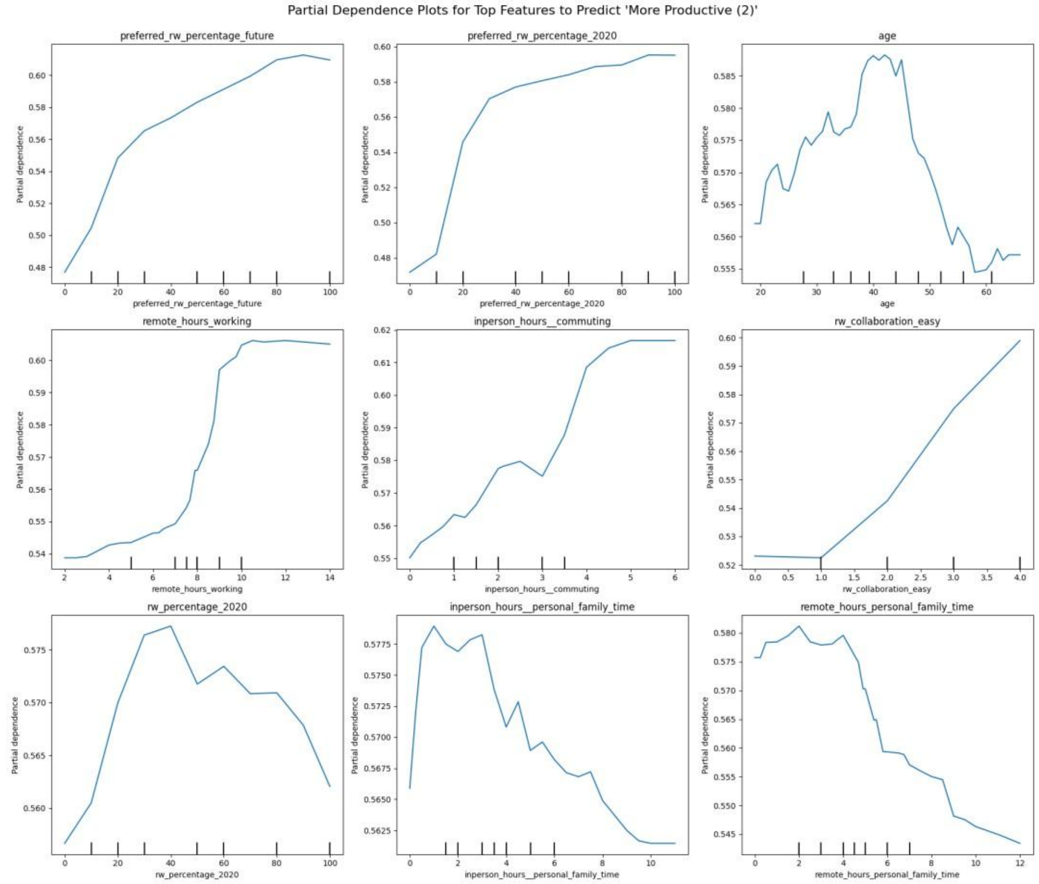
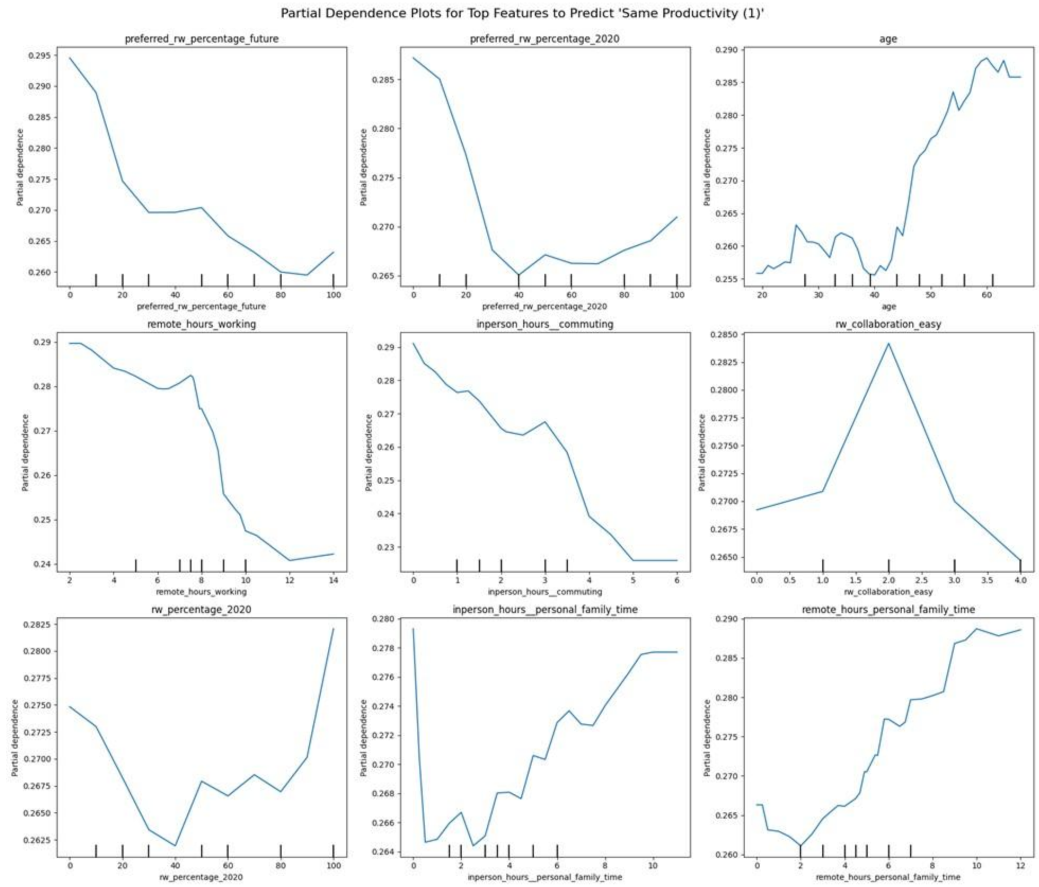


Figure 4 – *Partial Dependence Plots*





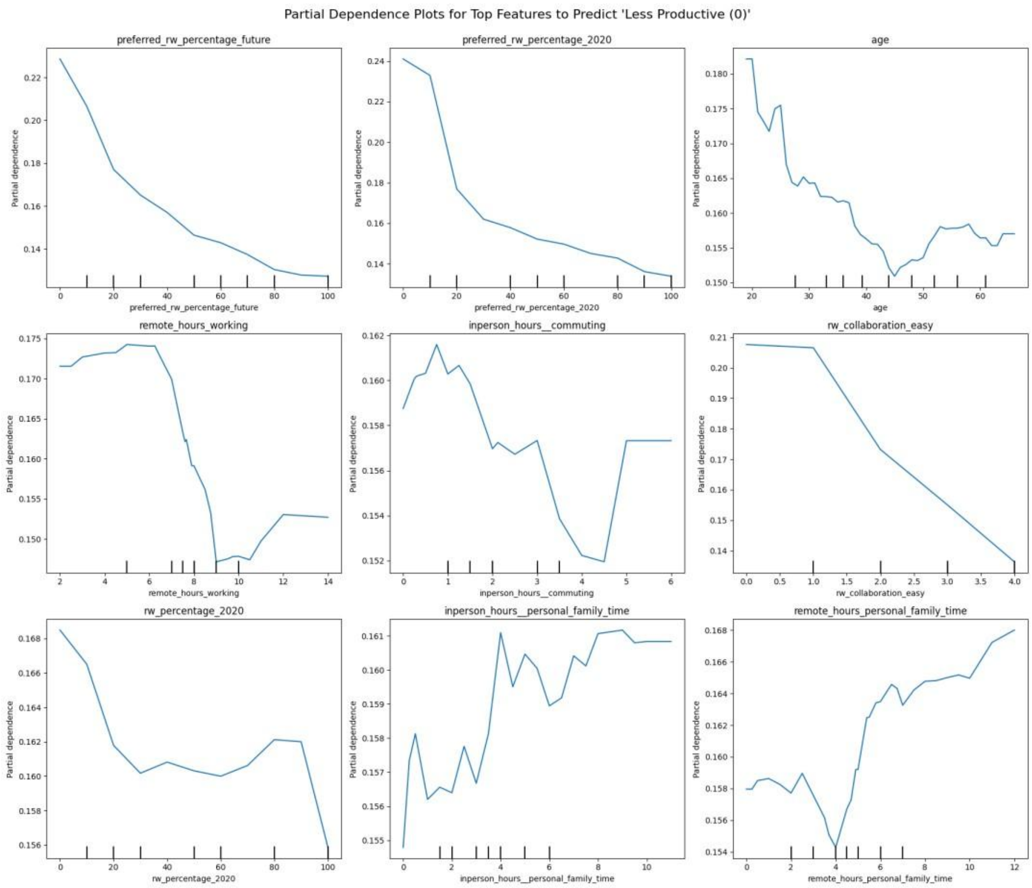
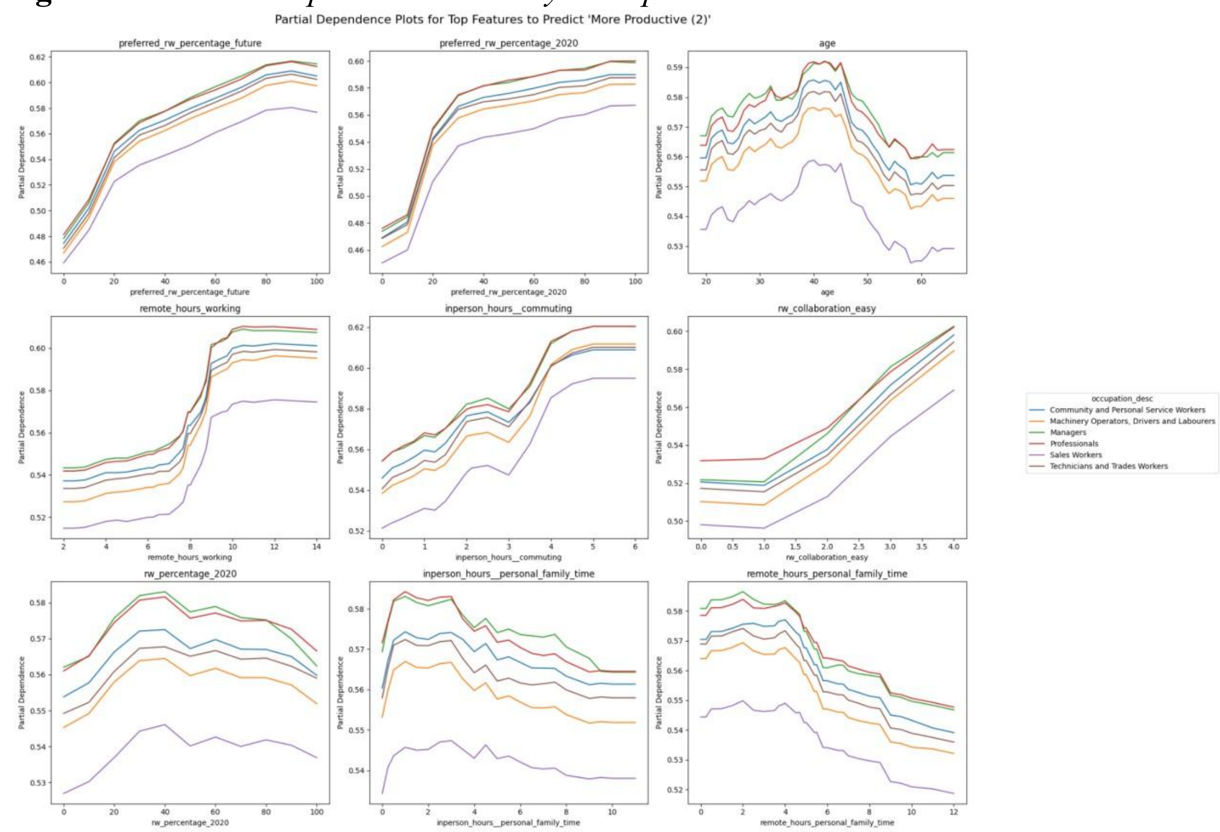
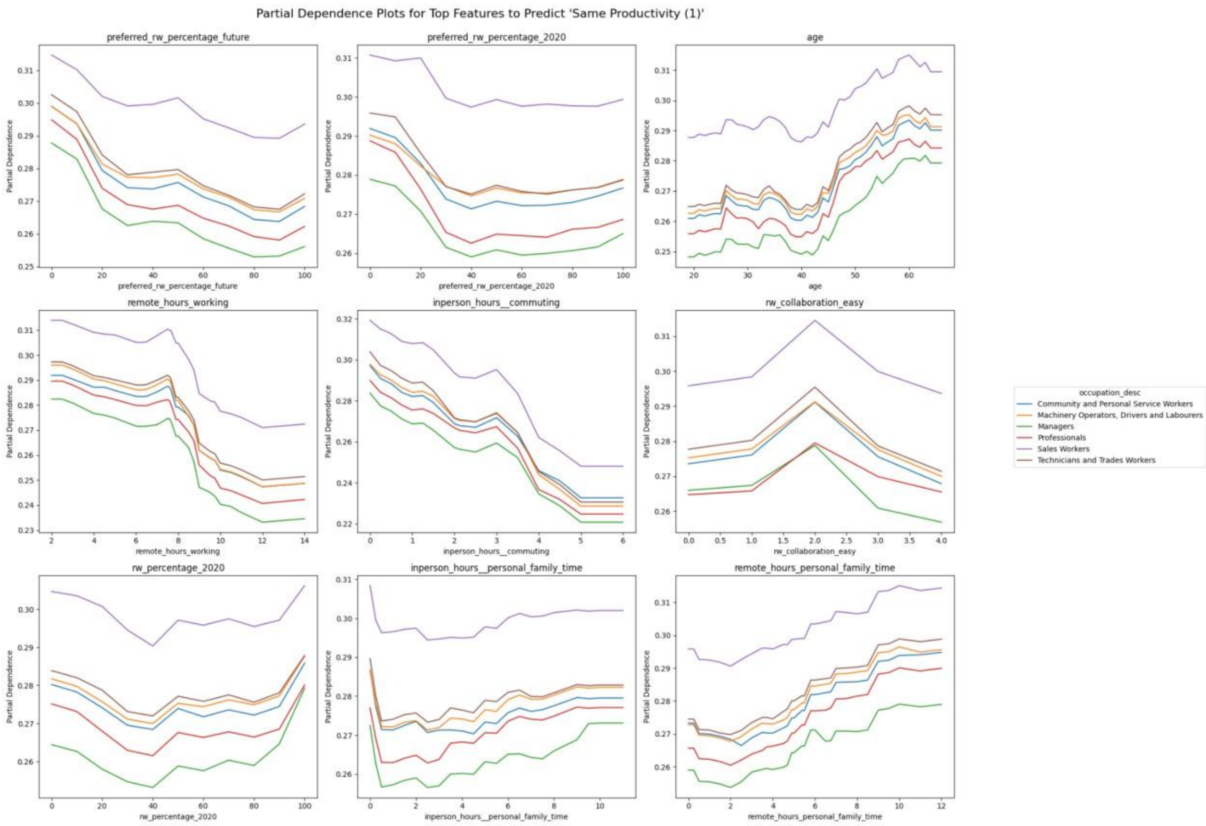
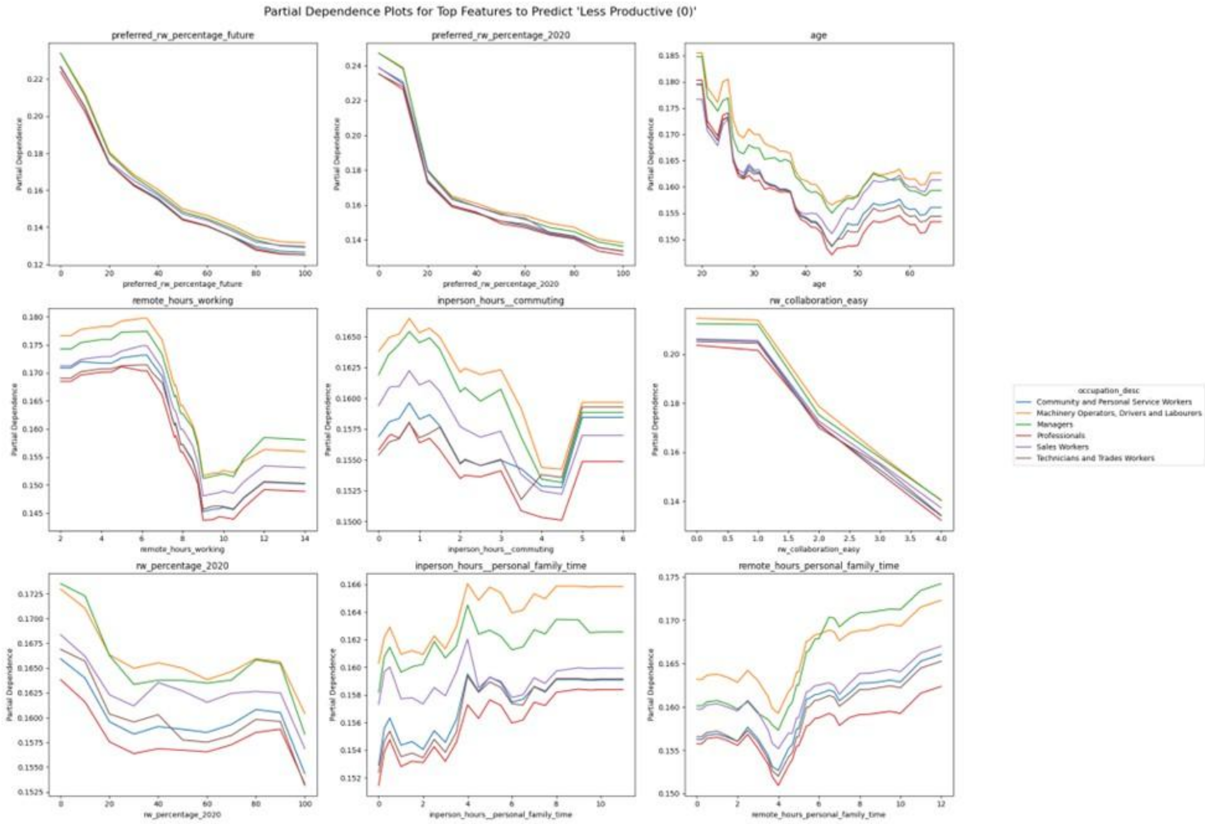


Figure 5 – *Partial Dependence Plots by Occupation*




Partial Dependence Plots for Top Features to Predict 'Same Productivity (1)'
preferred_rw_percentage_future
preferred_rw_percentage_2020
age
remote_hours_working
inperson_hours__commuting
rw_collaboration_easy
rw_percentage_2020
inperson_hours__personal_family_time
remote_hours_personal_family_time
Partial Dependence
occupation_desc
Community and Personal Service Workers
Machinery Operators, Drivers and Labourers
Managers
Professionals
Sales Workers
Technicians and Trades Workers


Partial Dependence Plots for Top Features to Predict 'Less Productive (0)'
preferred_rw_percentage_future
preferred_rw_percentage_2020
age
remote_hours_working
inperson_hours__commuting
rw_collaboration_easy
rw_percentage_2020
inperson_hours__personal_family_time
remote_hours_personal_family_time
Partial Dependence
occupation_desc
Community and Personal Service Workers
Machinery Operators, Drivers and Labourers
Managers
Professionals
Sales Workers
Technicians and Trades Workers