DOI 10.2478/sbe-2021-0008

SBE no. 16(1) 2021

USING NEURAL NETWORKS TO PREDICT WAGES BASED ON WORKER SKILLS

JAMES Otto

College of Business & Economics, Towson University, USA

HAN Chaodong

College of Business & Economics, Towson University, USA

TOMASI Stella

College of Business & Economics, Towson University, USA

Abstract:

This paper describes the results of our Neural Network (NN) models that predict annual wages based on the combination and levels of 35 different skills possessed by wage earners. These models can estimate the value of skills for skill-based compensation systems. They can be used by employers to determine how much to compensate different combinations of skills and by employees to estimate what they should be paid. Finally, governments can use the models to support the development and analysis of labor policies. We collect and integrate official U.S. Government data on 35 general job skills with the annual wage data for over 900 standard occupations. The skills data is then used as inputs to train an artificial intelligence (AI) neural network (NN) models. The resulting NN models train to above 70 percent accuracy in predicting annual wage levels based on the combination and levels of 35 different skills. This research makes use of authoritative U.S. Government data in new ways that can be used to better understand the connections between general skills and their relationships to wages. The need for these types of analytical tools is all the more important as changes in the job market have been severely impacted by the COVID 19 pandemic and are increasingly at risk of being replaced by automation.

Key words: *Neural Networks, Occupational Skills, Skill-based Pay, Salary Prediction*

1. Introduction

Currently, the COVID-19 pandemic has resulted in massive unemployment due to business closures and suppressed consumer demand (Tappe, 2020). For example, squeezed by debt and store closures, luxury retailer Neiman Marcus filed for chapter 11 during the COVID 19 pandemic (Friedman and Maheshwari, 2020). It is a distinct

probability that millions of workers may have to look for jobs across different industries when they return to the workforce.

While the pandemic will eventually end, industry experts cautioned that the historically low unemployment rates that existed before the pandemic will likely not return as many companies restructure or go out of business (Morath and Guilford, 2020). We note that meanwhile, the pandemic has affected different industries differently. For example, retail, restaurants, travel, and bars that require customers to congregate closely together may see permanently lower employment. It is a reality that millions of workers may have to look for jobs across different industries when they return to the workforce. To add to this unfortunate situation, Frey and Osborne (2017) suggested that “47% of total US employment is in the high risk category” of being replaced by technology in the next 10-20 years. Our research can help wage earners and employers move from unemployment to employment by defining the combinations of general skills that are most needed and what they are worth.

In our research, we analyze over 900 different combinations of up to 35 diverse occupational skills across multiple industries to assess their overall value, in terms of annual wages, in the job market. The predictive analytics leverage neural network (NN) models to provide valuable insights into how different combinations of skills impact wage levels. Our models could be especially useful because the unemployed may need to move from languishing industries to ones less impacted by COVID 19 or less threatened by automation.

Our research is organized as follows: we first review extant models used to predict employee salaries and then introduce our data collection process and present research data in detail, including dependent variables and independent variables in Section 3. Neural Network models are developed in Section 4, followed by results in Section 5. We conclude our research and discuss practical implications in Section 6.

2. Literature Review

Halteh, Arrowsmith, Parker, Zorn, and Bentley (2018) discussed a research agenda for employment research and stated one of the most urgent areas for research into technology from a business perspective includes implications for education, skills, and training. Our analysis extends current research by developing skill-based wage models that leverage the U.S. Government skills and salaries databases that we use to build our models. And none of the models use the types and number of predictor variables that we do (35) or the total number of jobs (937) as part of our model. This is a significant research gap that we are addressing in this paper.

Dilchert and Ones (2008) and Spurk and Abele (2011) predicted salary based on five individual personality dimensions based on psychological assessments. Nummi and Salonen (2005) used a Finnish dataset of 1998 retirees to build salary regression models based on yearly wages, working hours, date of birth, career duration, and sex.

In 2013, Kaggle Inc. (Kaggle, 2013) held a salary prediction competition using text-based job announcements where both job descriptions and salaries were included in the announcements. The Kaggle data set contained over 240,000 United Kingdom job

advertisements which included the job title, description, location, whether full-time or part-time, whether a permanent position or a limited duration contract, company name, job category, and annual salary. Jackman and Reid (2013) leveraged the Kaggle dataset to build and test models using a variety of regression methods, artificial neural networks, and random forests. Zhou and Liu (2016) used the same Kaggle dataset to build linear regression, Tikhonov Regression, Nearest Centroid Classifier, Nearest Neighbor Regression, Support Vector Machine (SVM), and Random Forest models. They were able to obtain a normalized mean absolute error (MAE) of 82% using random forest.

Recruiting firms have built models to estimate salaries (see Table 1) based on very specific inputs such as an individual's location, education, experience, and specific skillset. As can be seen in Table 1, these models accept very specific skills related only to very specific occupations, such as 'Apache Kafka' and 'multithreading Java programming'. The models in Table 1 also use detailed qualifications that are specific to individuals (such as years of experience). These skills and attributes are generally not relevant across a wide array of job types and industries. Contrast them with the more all-purpose skills that we incorporate in our models such as active listening or social perceptiveness. The full list of 35 skills that we include in our modeling is provided in Table 3.

Table 1: Salary Prediction Models

Salary Predictor	Sample Input Data	Reference
Dice	Job Title; Location; Years of Experience; Detailed Skills for Job Title (e.g., Multithreading, Apache Kafka, etc...)	Dice, 2019
PayScale	Job Title; Years in Field; City; Foreign Language; Skills Critical to Job Title; Type of employer; Work Venue	Payscale, 2019
Glassdoor	Current Employer; Employer Location; Job Title; Years of Relevant Experience; Base Salary; Additional Compensation; Highest Level of Education; University; Major/Concentration	Glassdoor, 2019
Salary	Job Title; Location; Education; Years of Experience; Direct Reports; Reports To; Performance	Salary, 2019
Indeed	Job Title; Location	Indeed, 2019
SalaryExpert	Job Title; Location; Highest Level of Education; Major/Field of Study; School Attended; Skills or Specialties; Current Employer Name; Years of Relevant Experience; Current Employer; Years of Experience; Salary	SalaryExpert, 2019
Adzuna	Title, Full Description; Location Raw; Location Normalized; Contract Type; Contract Time; Company; Category; Salary Raw; Salary Normalised; Source Name	Adzuna, 2019

For employees and job seekers, our models can be used to help individuals manage their skill inventories by predicting the wages of different skillset combinations. Individual workers can use the data to estimate what they should be paid based on the amalgam of skills they bring to the table.

This is important because workers must increasingly take responsibility for managing their careers and concerning themselves with maintaining and upgrading their skillsets to adapt to fast-changing work environments. Hall (1996) describes the protean career, which is driven by the individual rather than the organization. The protean career

requires the individual to personally acquire and use an identifiable set of skills. Laar, van Deursen, van Dijkstra, and Haan (2017) argue that the current workplace requires highly skilled workers who must address increasingly complex and interactive tasks. Employees not only need excellent technical preparation; they also need sufficient skills to adapt to the changing requirements of the job (Carnevale and Smith, 2013). For individuals, our models can provide wage information to help them formulate which skill combinations to pursue.

Our models leverage free, publicly available U.S. government data. Since the data comes from the U.S. Federal government, it may have a higher level of intrinsic credibility. Moreover, since the data collection and maintenance costs are borne by the U.S. government rather than by a specific company, this may incentivize individuals and companies to use this free data to improve their analytics. The government data uses standardized occupational job codes, so the data can be easily associated and integrated across the federal government. This makes analysis across multiple government datasets quicker and simpler.

3. Data

We collect occupation skills and wage data sponsored by the U.S. Government as inputs to develop, train, and test our NN models that are described in this section. Specifically, the wage data is provided by the U.S. Bureau of Labor Statistics (BLS) and covers over 900 different job categories. The occupation skills data, including Skill Level Ratings and their related Skill Importance Ratings, are collected from the Occupational Information Network (O*NET), which is sponsored by the U.S. Department of Labor Employment and Training Administration. Updated annually, the O*NET database contains hundreds of standardized and occupation-specific descriptors on nearly a thousand job occupations covering the entire U.S. economy. It is considered the primary source of U.S. occupational information (O*NET, 2019). All 35 different occupational skills organized into six categories by O*NET are listed in Table 3.

3.1 Dependent Variable - Wages

The dependent variable output of the models are estimates of the wages for an employee with different combinations and levels of skills. The annual wage data used for training and testing the models is provided by the BLS Occupational Employment Statistics (OES). The BLS is the principal federal agency responsible for measuring labor market activity, working conditions, and price changes in the economy (BLS, 2019). The BLS OES provides nationwide, average annual wage data for close to 1000 Occupations in the U.S. across numerous industries and job types. Their dataset includes wage data from high-level positions, such as Chief Executives, to low wage employees, such as fast-food workers.

The BLS OES data is gathered from program semi-annual surveys to estimate wages by national, state, and metropolitan areas. We used the 2017 national data. Wages for the BLS OES survey are defined as straight-time, gross pay, which is exclusive of premium pay. The survey covers all full-time and part-time wage and salary workers in

nonfarm industries. The wage data includes base rate, cost-of-living allowances, guaranteed pay, hazardous duty pay, incentive pay, on-call pay, and tips. The wage data does not include back pay, jury duty pay, overtime pay, severance pay, shift differentials, nonproduction bonuses, tuition reimbursements, or data from self-employed workers. The BLS OES surveys approximately 180,000 to 200,000 establishments every six months - so it takes three years to completely collect the full sample of 1.2 million establishments (BLS, 2019).

The BLS occupational annual wage data was integrated with the O*NET occupational skills data via job codes to create a consolidated dataset that links the 35 skillsets to the wage data for 967 occupations. The initial dataset was then pruned to 937 occupations to eliminate missing and incomplete data. In addition, we assigned a level of zero (0) for skill level ratings or skill importance ratings with 'Not available' or 'Not relevant' as their value. Summary statistics for the data are provided in Table 2.

Table 2: Summary Statistics of Annual Pay (\$)

Mean	Median	Mode	St Dev.	Kurtosis
65,019	56,510	91,080	36,702	6
Skewness	Range	Min	Max	
2	244,760	21,230	265,990	

According to Table 2, while the average annual pay was \$65,019 for all 937 occupations, the range is enormous with anesthesiologists averaging \$265,990 and food preparation and serving workers averaging only \$21,230. Note that the 21 highest paid occupations are highly specialized medical professionals (job category coded 29) and CEOs averaged \$196,050, ranking 22nd. The distribution of annual pay for all occupations is presented in Figure 1.

Figure 1: Histogram of Annual Pay

Based on Figure 1, 42% of the occupations (394) earn less than \$50,000 in annual pay and 69% of the occupations make less than \$75,000 in annual pay. Less than 3% of the occupations (24) make more than \$200,000 a year.

3.2 Independent Variables

The independent variables used for analysis in this research are the Skill Level Ratings and the Skill Importance Ratings of all 35 skills (see Table 3) that are associated with 937 occupations and range from 0 to 100. Such data are collected from the O*NET dataset, which contains several hundred descriptive ratings from O*NET questionnaire responses from sampled workers, occupational experts, and occupation analysts (Fleisher and Tsacoumis, 2012).

Randomly sampled workers answer one of several possible 30-minute questionnaires. For occupations where it would be difficult to sample workers, such as those in remote locations, occupation experts are identified and sampled from professional and/or trade association membership lists. The occupation experts complete multiple questionnaires. In addition to the questionnaires completed by workers and occupation experts, additional ratings are provided by occupation analysts. Responses from all three sources (workers, occupation experts, and occupation analysts) are used to provide complete information for each occupation.

Table 3: Occupational Skills

Skill Category	Skill Description	Skills
Basic Skills (10)	Developed capacities that facilitate learning or the more rapid acquisition of knowledge	<i>Active Learning, Active Listening, Critical Thinking, Learning Strategies, Mathematics, Monitoring, Reading Comprehension, Science, Speaking, Writing</i>
Social Skills (6)	Developed capacities used to work with people to achieve goals	<i>Coordination, Instructing, Negotiation, Persuasion, Service Orientation, Social Perceptiveness</i>
Technical Skills (11)	Developed capacities used to design, set-up, operate, and correct malfunctions involving the application of machines or technological systems	<i>Equipment Maintenance, Equipment Selection, Installation, Operation and Control, Operation Monitoring, Operations Analysis, Programming, Quality Control Analysis, Repairing, Technology Design, Troubleshooting</i>
Complex Problem-Solving Skills (1)	Developed capacities used to solve novel, ill-defined problems in complex, real-world settings	<i>Complex Problem-Solving</i>
Systems Skills (3)	Developed capacities used to understand, monitor, and improve socio-technical systems	<i>Judgment & Decision Making, Systems Analysis, Systems Evaluation</i>
Resource Management Skills (4)	Developed capacities used to allocate resources efficiently	<i>Management of Financial Resources, Management of Material Resources, Management of Personnel Resources, Time Management</i>

Each skill in Table 3 is described along with descriptive anchors for low, medium, and high skill level ratings (which range from 0 to 100). These anchors help the analysts determine the skill levels associated with the different occupations. For the O*NET data, the low anchor value for all skills is 28, the medium anchor value is 57, and the high value anchor is 85. For example, the definitions for the *Active Learning* and the *Complex Problem Solving* skills, along with their anchor descriptions, are summarized below.

- **Active Learning:** Understanding the implications of new information for both current and future problem-solving and decision-making.
 - **Low Anchor (27):** Think about the implications of a newspaper article for job opportunities
 - **Medium Anchor (57):** Determine the impact of new menu changes on a restaurant's purchasing requirements
 - **High Anchor (85):** Identify the implications of a new scientific theory for product design
- **Complex Problem Solving:** Identifying complex problems and reviewing related information to develop and evaluate options and implement solutions.
 - **Low Anchor (27):** Lay out tools to complete a job
 - **Medium Anchor (57):** Redesign a floor layout to take advantage of new manufacturing techniques
 - **High Anchor (85):** Develop and implement a plan to provide emergency relief for a major metropolitan area

In addition to the Skill Level Ratings, the related importance of each skill (Skill Importance Rating) to each occupation is also provided, also ranging from 0 to 100 (O*NET, 2019), using the following anchors:

0 = Not Important
 25 = Somewhat Important
 50 = Important
 75 = Very Important
 100 = Extremely Important

For example, the *Marketing Manager* occupation requires mathematic skills at a Skill Level Rating of 45 out of 100, with a Skill Importance Rating of 44 out of 100. Contrast that occupation with the *Refuse and Recyclable Material Collectors* occupation, which requires a *mathematics* Skill Level Rating of 7 out of 100, with an Importance Rating of 10 out of 100. Thus, the level of *mathematics* skills required of *Marketing Managers* is much higher and significantly more important for them, than for *Refuse and Recyclable Material Collectors*.

Summary statistics for the Skill Level ratings and the Skill Importance ratings are presented in Tables 4 and 5 respectively.

As shown in Table 4, the overall average level rating across all occupations is 34 with great variation across skills. High overall level ratings indicate that a high level of a particular skill may be demanded across the board or by a majority of occupations, while low overall level ratings suggest otherwise. For example, the *Reading Comprehension* skill has the highest overall level rating of 52, ranging from 23 to 86 across occupations; the *Critical Thinking* skill has the second-highest overall level rating of 51, ranging from 27 to 80 across occupations; and the *Active Listening* skill has the third-highest overall level rating of 50, ranging from 29 to 84 across occupations. High overall level ratings suggest that a high level of *Reading Comprehension*, *Critical Thinking*, and *Active Listening* skills

are desired across the board. Not surprisingly, the *Installation* skill has the lowest overall level rating of 4, ranging from 0 to 52 across occupations.

Some other highly technical and specialized skills are associated with low overall level ratings, including *Programming*, *Technology Design*, *Repairing*, *Equipment Selection*, and *Equipment Maintenance*. The O*NET dataset shows that only a small number of occupations desire a high level of such skills and a vast majority of other occupations do not need them at all. For example, 776 out of 937 occupations do not even need any level of *Installation* skill (0 level rating).

According to Table 5, the overall average importance rating across all occupations is 40 with great variation across skills. High overall importance ratings indicate that a particular skill may be considered very or extremely important across the board or by a majority of occupations while low overall import ratings suggest otherwise. For example, the *Active Listening* skill has the highest overall importance rating of 65, ranging from 25 to 97 across occupations; the *Speaking* skill has the second-highest overall importance rating of 64, ranging from 28 to 94 across occupations. It is no surprise that *Active Listening* and *Speaking* are considered very important or extremely important skills for the success of most occupations. In contrast, the *Installation* skill has the lowest overall importance rating of 6, ranging from 0 to 72 across all occupations; the *Programming* skill has the second-lowest overall importance rating of 15, ranging from 0 to 94 across all occupations. A significant number of occupations do not consider *Installation*, *Programming*, *Repairing*, *Technology Design*, *Equipment Selection*, or *Equipment Maintenance* skills important at all. For example, many of the 937 occupations do not consider the *Installation* skill important at all, with zero (0) importance level ratings.

Our analysis examines different ways of using occupations' Skill Level Ratings and their associated Skill Importance Ratings to assess their impact on wages. That is:

- **Skill Level Ratings Only:** Only the Skill Level Rating data are used as inputs to the models. The associated Skill Importance Rating data are not used.
- **Skill Level Ratings Only (When Skill Importance > 50):** Occupational Skill Level Ratings are only used if their associated Skill Importance Ratings are greater than 50 (out of 100). The associated Skill Importance Rating data are not used.
- **Skill Level * Skill Importance:** The Skill Level Ratings are multiplied by the Skill Importance Ratings for each occupation.

Table 4: Summary Statistics for the Level Ratings of Skills

	Mean	Median	Mode	St. Dev.	Min	Max
Management of Material Resources	21	20	0	12	0	68
Management of Personnel Resources	37	37	43	10	0	77
Management of Financial Resources	19	16	0	14	0	79
Time Management	44	43	43	7	20	68
Speaking	49	50	57	10	27	77
Science	20	13	0	21	0	84
Reading Comprehension	52	54	57	12	23	86
Monitoring	48	48	43	8	0	79
Mathematics	37	36	30	14	0	87
Learning Strategies	40	41	43	11	0	75
Critical Thinking	51	52	57	9	27	80
Active Learning	46	45	43	11	16	80
Writing	47	46	57	12	7	82
Active Listening	50	54	57	9	29	84
Complex Problem Solving	45	45	43	10	0	75
Social Perceptiveness	44	43	43	10	23	84
Service Orientation	41	41	43	9	13	71
Persuasion	40	41	43	10	0	71
Negotiation	37	37	41	10	0	71
Instructing	42	43	43	11	0	70
Coordination	45	45	43	7	0	73
Systems Evaluation	37	39	43	13	0	73
Systems Analysis	38	39	43	12	0	77
Judgment and Decision Making	46	45	43	9	23	82
Equipment Maintenance	15	0	0	18	0	68
Installation	4	0	0	10	0	52
Operation and Control	25	29	0	18	0	80
Operation Monitoring	33	34	45	14	0	70
Quality Control Analysis	30	34	0	16	0	61
Operations Analysis	22	20	0	18	0	71
Programming	10	9	0	12	0	70
Troubleshooting	22	23	0	17	0	57
Repairing	14	0	0	18	0	61
Technology Design	13	11	0	11	0	59
Equipment Selection	14	9	0	16	0	59
Average over all skills	34	33	29	12	6	73

Table 5: Summary Statistics for the Importance Ratings of Skills

	Mean	Median	Mode	St. Dev.	Min	Max
Management of Material Resources	25	25	25	11	0	72
Management of Personnel Resources	41	41	47	13	0	81
Management of Financial Resources	22	19	19	12	0	78
Time Management	53	50	50	8	25	75
Speaking	64	66	75	12	28	94
Science	25	19	0	23	0	97
Reading Comprehension	62	66	75	14	22	97
Monitoring	58	56	50	9	31	85
Mathematics	39	38	25	15	0	100
Learning Strategies	44	44	50	14	3	85
Critical Thinking	63	66	75	11	31	94
Active Learning	53	53	50	12	22	78
Writing	54	53	53	15	10	97
Active Listening	65	69	75	11	25	97
Complex Problem Solving	55	53	50	12	22	85
Social Perceptiveness	54	50	50	11	25	94
Service Orientation	49	47	50	13	10	81
Persuasion	44	44	47	12	16	81
Negotiation	41	41	47	12	13	91
Instructing	46	47	50	14	3	91
Coordination	53	50	50	9	25	81
Systems Evaluation	41	44	25	14	3	78
Systems Analysis	43	44	50	15	0	78
Judgment and Decision Making	57	56	50	11	25	88
Equipment Maintenance	18	3	0	21	0	81
Installation	6	0	0	13	0	72
Operation and Control	30	28	0	22	0	97
Operation Monitoring	39	38	22	19	0	94
Quality Control Analysis	36	38	50	17	0	78
Operations Analysis	27	25	22	16	0	72
Programming	15	13	0	13	0	94
Troubleshooting	26	25	0	19	0	81
Repairing	17	3	0	22	0	81
Technology Design	18	19	19	10	0	66
Equipment Selection	18	13	0	17	0	69
Average over all skills	40	38	36	14	10	85

4. Development of Neural Networks Model

Artificial neural networks are a type of machine learning, a subset of artificial intelligence (AI), that provide information systems with the ability to learn automatically from example data without human intervention. As computers learn directly from data, the need for humans to specifically program is reduced. This is a substantial benefit because the computer can automatically generate models independent of explicit human programming, which can be time-consuming and costly. This machine learning cost

advantage, combined with large amounts of freely available government data (BLS, 2019; ONET, 2019), provides an inexpensive way to develop analytical tools like the ones described in this paper.

We used the open-source RStudio integrated development environment, along with the R programming language to build several NN models from the occupational and wage data. There are over 25 different NN architectures available to choose (Mehta, 2019). We selected the backpropagation NN architecture because it is a widely used, classical architecture that is suitable for the type of analysis we are performing. The model development steps are summarized below.

1. Integrate 35 occupational skills data (independent variables) with associated national wage data (dependent variable)
2. Build, train, test, and run backpropagation NNs (using 35 occupational skills and their wages)
 - The NN structure hyperparameters consisted of 35 input nodes, with a three node hidden layer, another single node hidden layer, and a single output node. Two hidden layers were chosen because a two hidden layer architecture can approximate any smooth mapping to any accuracy (Heaton, 2017). Training time was not a significant issue given the reasonable size of our dataset. Several models were tried using different structure hyperparameters. The architecture with a 3-node hidden layer and another single node hidden layer provided performance similar to more complex NNs with larger numbers of nodes in the hidden layers.
 - The default logistic activation function is used in the nodes. The main reason why we use logistic function is that it produces a continuous, as opposed to discrete, outputs between 0 and 1.

5. Results

The results of running the models are provided in Table 6, which provides accuracy for the NN models. The table provides the averages of running the NN models ten times with random training/testing using 90/10 data splits.

Table 6: Neural Network Results

	Mean Accuracy
Skill Levels * Skill Importance	0.764
Skill Levels Only	0.741
Skill Levels Only (Importance > 50)	0.710

Mean Accuracy is defined as $1 - \text{mean}\left(\frac{|A-P|}{A}\right)$, where A is the actual salary, and P is the predicted salary. Thus, table 6 shows us that for the *Skill Levels Only* inputs (Mean Accuracy = 0.741), on average the difference between the predicted and actual salaries will be less than 26 percent. For example, this means that our prediction for a job with a \$50K salary would on average, fall between \$37K and \$67K.

Figure 2 demonstrates this visually. It compares the predicted wage values against the actual wage data for the NN model. A perfect fit between actual and predicted salaries would show the plot points falling perfectly on the line for a mean square error (MSE) of zero.

Figure 2: NN Predicted vs. Actual Wages

Thus, for instance, we can see in Figure 2 that salary predictions for jobs paying below \$50K per year have lower error (the data points are packed closer to the zero-error line) than jobs that pay above \$100K per year (where the data points are for the most part farther from the zero-error line).

6. Conclusions and Implications

We have created wage prediction models that use freely available standardized skills and wage data that can be used by organizations and individuals to determine the wage values of different combinations of general skill sets. Since the data is collected and maintained by the Federal government, this may provide a relatively low-cost way for organizations and individuals to conduct salary studies based on worker skills.

These models are appropriate for the analysis of wages based on general skillsets (see Table 3), that apply across nearly all occupations and industries rather than the types of very specific occupation-related skills listed in Table 1. These more general models may be better suited for individuals, such as students, who may have these more general skills, rather than the work experience or occupation-specific skill sets of Table 1.

An employee with a certain set of skills can look to the closest jobs across industries by comparing the required skillset and expected salaries. For example, the closest job for *Pump Operators, Except Wellhead Pumpers* (coded 53-7072.00) in

Transportation and Material Moving Occupations to move up is *Hydroelectric Production Managers*, (coded 11-3051.06) in Management Occupations.

For the government, our tools can be used to help develop labor policies to improve those skills that are the most important to the labor marketplace, as measured by wages. For employing organizations, our model results can help HR departments determine what different skill sets should cost in wages, and budget their pay scales accordingly. These insights can support the management of pertinent employee skills needed to keep apace and stay relevant to the marketplace. Our models can provide insights to help organizations make decisions about how to best manage critical skill needs driven by the rapid, and constant, advance of technology.

Since the data comes from the U.S. Federal government, it may have a higher level of intrinsic credibility. And since the data collection and maintenance costs are borne by the U.S. government rather than by a specific company, this may incentivize individuals and companies to use this free data to improve their analytics. The government data uses standardized occupational job codes, so the data can be easily associated and integrated across the federal government. This makes analysis across multiple government datasets quicker and simpler.

7. References

- Adzuna (2013), *Job Salary Prediction*, available online at <https://www.kaggle.com/c/job-salary-prediction/overview/timeline>.
- Bureau of Labor Statistics (2019), *Occupational Employment Statistics Overview*, Available online at https://www.bls.gov/oes/oes_emp.htm.
- Carnevale, A., and Smith, N., (2013), Workplace basics: the skills employees need and employers want, *Human Resource Development International*, Vol. 16, No. 5.
- Dice.com (2019), *How Much Are Your Skills Worth?* Available online at <https://www.dice.com/salary-calculator>.
- Dilchert, S. and Ones, D., (2008), Predicting managerial salary at different organizational levels, *Journal of Personal Psychology*, Vol. 7, pp. 1-23.
- Fleisher, M. and Tsacoumis, S., (2012), O*NET analyst occupational skills; ratings: procedures update, *Report FR-11-67*.
- Friedman, V., and Maheshwari, S., (2020), Neiman Marcus, a symbol of luxury, files for bankruptcy, *New York Times*, Available online at <https://www.nytimes.com/2020/05/07/business/neiman-marcus-bankruptcy.html>.
- Frey, C. and Osborne, M., (2017), The future of employment: how susceptible are jobs to computerisation? *Technological Forecasting & Social Change*, Vol. 114, pp. 254–280.
- Glassdoor.com., (2019), *Know Your Worth*, Available online at <https://www.glassdoor.com/Salaries/know-your-worth.htm>.
- Hall, J., (1996), Protean careers of the 21st century, *Academy of Management Executive*, 10 (4).
- Halteh, J., Arrowsmith, J., Parker, J., Zorn, T. E., & Bentley, T., (2018), The impact of technology on employment: a research agenda for New Zealand and beyond, *Labour & Industry*, 28 (3), 203–216.
- Heaton, J., (2017), *The Number of Hidden Layers*, Available online at <https://www.heatonresearch.com/2017/06/01/hidden-layers.html>.
- Indeed.com., (2019), *Search and Compare Salaries*, Available online at <https://www.indeed.com/salaries>.

- Jackman, S. and Reid, G., (2013), Predicting job salaries from text descriptions, Ph.D. dissertation, University of British Columbia, Available online at <https://open.library.ubc.ca/cIRcle/collections/graduateresearch/42591/items/1.0075767>.
- Kaggle Inc., (2013), *Job Salary Prediction*, Available online at <http://www.kaggle.com/c/job-salary-prediction>.
- Laar, E., van Deursen, A., van Dijka, J., and Haan, J., (2017), The relation between 21st-century skills and digital skills: a systematic literature review, *Computers in Human Behavior*, Vol. 72, pp. 577-588.
- Mehta, A., (2019), A comprehensive guide to types of neural networks, *Digital Vidya Data Analytics BLOG*, Available online at <https://www.digitalvidya.com/blog/types-of-neural-networks/>.
- Morath, E., and Guilford, G., (2020), Another 3.2 million workers filed for jobless benefits last week, *Wall Street Journal*, Available online at https://www.wsj.com/articles/unemployment-benefits-weekly-jobless-claims-coronavirus-05-07-2020-11588813872?mod=article_inline&mod=hp_lead_pos1.
- Nummi, T. and Salonen, J., (2005), *Modeling and Predicting Individual Salaries: A Study of Finland's Unique Dataset*, Available online at <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.492.1881&rep=rep1&type=pdf>.
- O*NET (2019), *About O*NET*, Available online at <https://www.onetcenter.org/overview.html>.
- Payscale.com., (2019), *Pay Report*, Available online at <https://www.payscale.com/my/survey/job>.
- Salary.com., (2019), *What Am I Worth?* Available online at <https://www.salary.com/>.
- SalaryExpert.com., (2019), *Calculate your Salary*, Available online at <https://www.salaryexpert.com/salarycalculator>.
- Spurk, D. and Abele, A., (2011), Who earns more and why? A multiple mediation model from personality to salary, *Journal of Business Psychology*, Vol. 26, pp. 87–103.
- Tappe, A., (2020), "April was probably the worst month for American jobs since the great depression", *CNN Business*, Available online at <https://www.cnn.com/2020/05/06/economy/april-jobs-report-2020/index.html>.
- Zhou, Y., Li, L., and Liu, X., (2016), Prediction of salary in the UK, *Working Papers, University of California, San Diego*, Available online at <http://jmcauley.ucsd.edu/cse190/projects/fa15/012.pdf>.