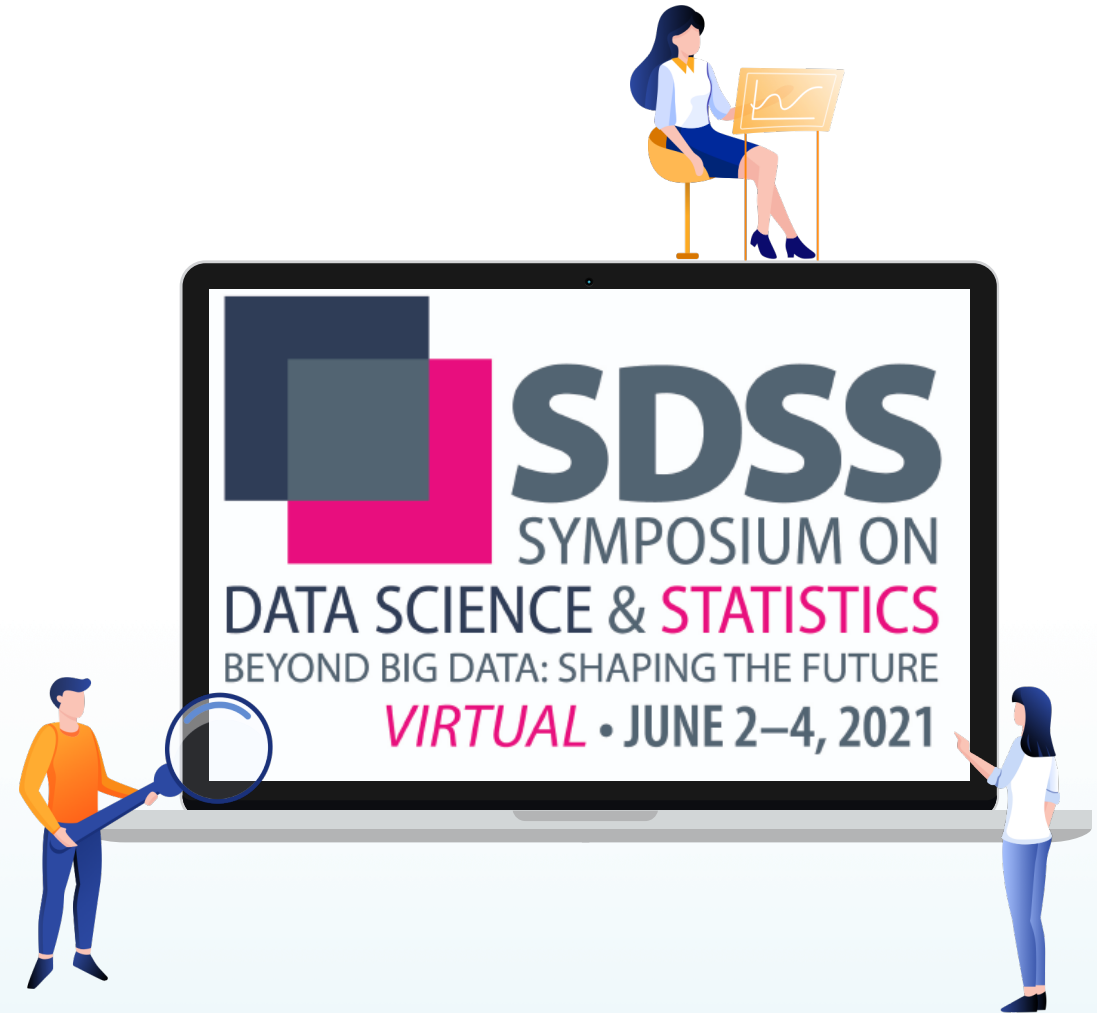


Auto-Classification of Occupational Data

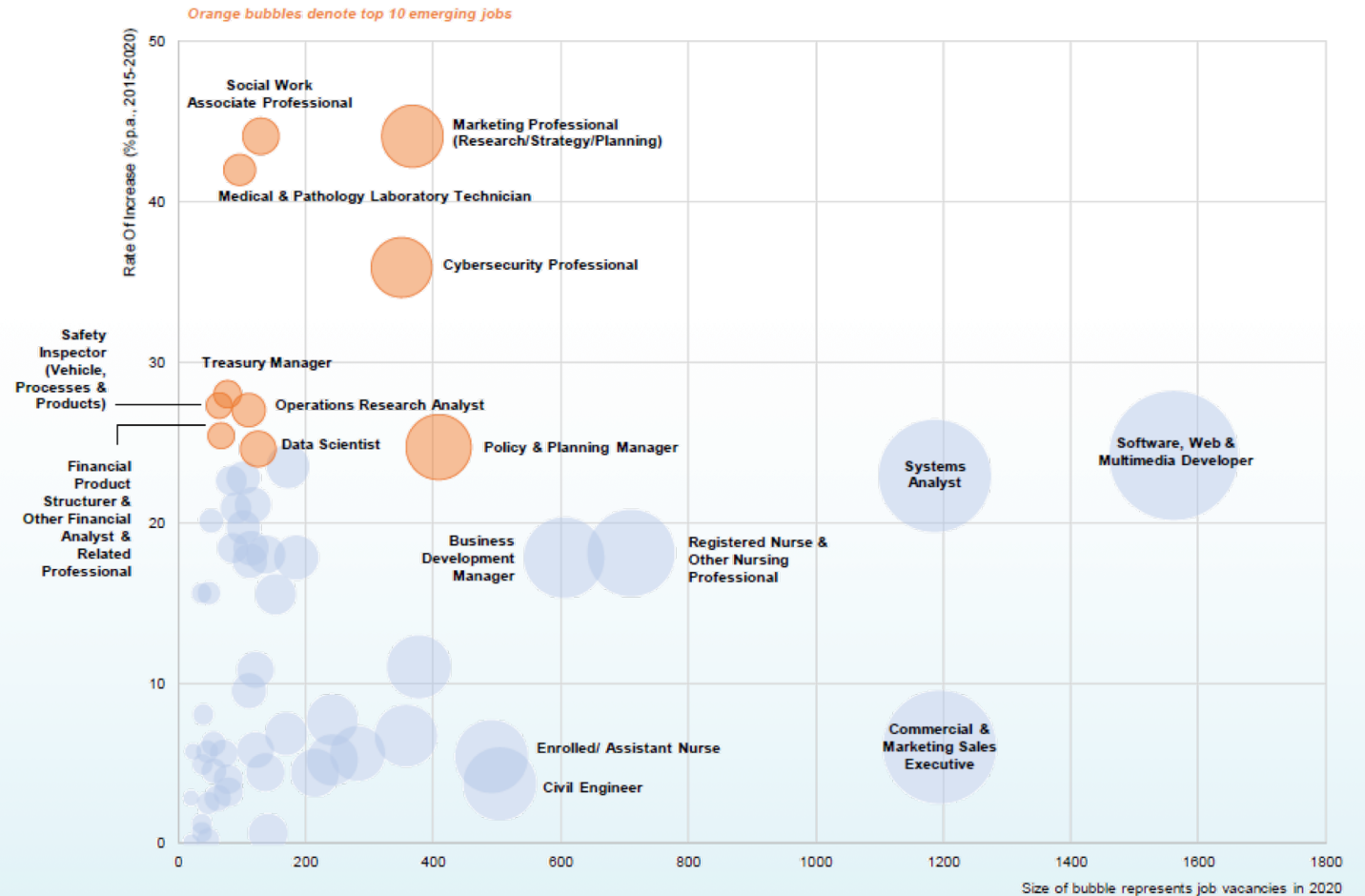
Presented by Chong Ning
Ministry of Manpower, Singapore
3rd June 2021, 10:00 AM to 11:35 AM



Background Information

- The Ministry provides information and statistics on Singapore's workforce and labour market.
- Occupational data is highly sought after as it provides a more detailed view on the workforce.
- Such data are used in reports about Job Vacancy and Skills Mismatches.

Figure 7: PMET Job Vacancies Which Saw An Increase In Demand From 2015 To 2020



Problem Statement

Occupational data is **complex** and **tedious** to collect.

Interviewers have to **manually classify** the occupation of an individual into one of thousands of occupational codes.



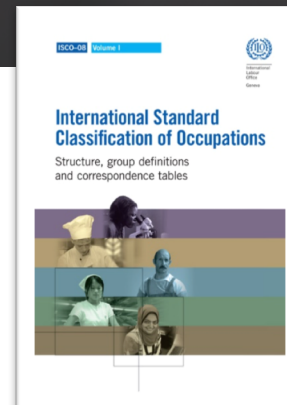
Occupational Data

- Occupational data is collected in an open ended format so that more information can be obtained.
- We ask the following questions to the respondent:
 - What is your job title?
 - What is your main task and duties?



Occupational Data

- The **International Standard Classification of Occupations (ISCO)** is a tool for categorizing jobs into a defined set of groups.
- **MRSD** adopts the **Singapore Standard Occupation Classification (SSOC)**.



Major Group	
1	Legislators, Senior Officials and Managers
2	Professionals
3	Associate Professionals and Technicians
4	Clerical Support Workers
5	Service and Sales Workers
6	Agricultural and Fishery Workers
7	Craftsmen and Related Trades Workers
8	Plant and Machine Operators and Assemblers
9	Cleaners, Labourers and Related Workers
X	Workers Not Elsewhere Classified

Occupational Data

Occupation 23101 University lecturer

University lecturer prepares and delivers lectures while conducting tutorials, seminars and laboratory experiments in one or more subjects within a prescribed course of study at a university. He/she supervises and grades experimental, practical work and examinations undertaken by students in addition to conducting academic research and publishing his/her findings in his/her respective field.

Occupation 23102 Polytechnic lecturer

Polytechnic lecturer prepares and delivers lectures while conducting tutorials, seminars and laboratory experiments in one or more subjects within a prescribed course of study at a polytechnic. He/she supervises and grades experimental, practical work and examinations undertaken by students in addition to conducting academic research and publishing his/her findings in his/her respective field.



**Singapore Standard Occupational
Classification 2020**

Example of Responses Collected

Responses from the ground

Category	Job Title	Job Desc	
22200	nurse	nursing	10
	senior staff nurse	nursing	8
23400	teacher	teach mathematics in secondary school	8
23500	teacher	teach chinese in primary school	9
		teach english and maths in primary school	9
24111	accountant	accounting	17
25160	project manager	direct and supervise information technology (it) plan..	10
33130	accounts executive	accounting	8
42245	customer service officer	customer service	10
51200	cook	cooking	10
52202	shop sales assistant	shop sales assistant duties	17
52302	cashier	cashiering	15
		cashiering duties	22
54142	security officer	security	17
		security services	11
83221	taxi driver	drive passengers	11
		drive passengers around for a fee	8
		drive taxi	28
		driver	13
		driving	13
		driving taxi	15
		ferry passengers	19
		ferrying passengers	13
83226	driver	driving	9
	grab driver	driver	20
		ferry passengers	12
91130	cleaner	clean office	9
		cleaning	19
		cleaning office	8
		ensure the cleanliness of the premises.	10
		office cleaner	9
		office cleaner	8
91190	cleaner	cleaning	39
		cleaning duties	11
		ensure the cleanliness of the premises.	16

Responses example - Lecturer

Category	Job Title	Job Desc	
23101	professor	professor. research and teach business strategy	1
		teaching	1
		teaching and research	4
		teaching organisational behaviour for phd programmes and executive education. resea..	1
		teaching research	1
	senior lecturer	teaching. research & development	1
		academic teaching. program management	1
		teach language	1
		teach mathematics to student teachers in degree programme and students in masters ..	1
		teach students in nus	1
23102	university lecturer	teach university students	1
		teacher	1
		teaching	1
		teaching accountancy in university	1
		prepare teaching materials. assessments and given lecture	1
	polytechnic lecturer	teacher at university	1
		university level lecture	1
		give class in nyp	1
		gives lectures to students in a polytechnic	1
		giving lectures and teaching students in the polytechnic	1
	senior lecturer	lecturer	1
		develop curriculum and deliver part-time and full-time courses for students and adult l..	1
		duties of lecturer. conduct lectures. manage students. impart knowledge. manages sta..	1
		educator. motivator	1
		facilitate classes in diploma in business information systems (dbis)..create and update ..	1
		facilitate classes in diploma of infocomm security management (dism)..create and upd..	1
		lecture and tutor human resource management and fundamentals of management	1
		lecturer	2
		teach	1
		teach 3d animation and modelling for virtual reality in polytechnic	1

Document Term Matrix

Document Term Matrix Example

Category	admin	assistant	cleaner	director	driver	engineer	executive	Words manager	officer	operate	sales	senior	service	teacher	technician
11		30		633		7	104	184	96	35	7	41	4		
12	60	143	1	549	1	30	78	1,527	62	52	329	279	78		1
13	3	51	1	316	4	69	38	696	40	54	6	113	43	2	4
14	3	23		231		1	14	194	9	11	51	9	18	1	1
21	3	52		34		1,051	52	150	35	10	4	257	35		9
22		20		9			1	26	20	1		116			2
23		18		3		1	5	13	113			79	2	654	1
24	41	137		130		42	198	557	114	11	149	246	28		6
25	3	48		33		228	42	318	13	15	1	196	8		2
26		22		30		2	14	19	10			38	3		
29		1					2	2	2	1		3		1	
31	1	126		7		481	96	135	164	47	3	187	65		527
32	1	51		2		3	5	4	13	1	1	28	5		49
33	173	421		162	1	45	1,391	1,123	438	188	542	449	120	2	3
34		15		16	1	1	23	30	16	1	1	18	3	7	1
35	2	14		1		64	20	18	17	14	2	22	15		58
36		28	1	8		1	4	6	3	1		11	1	447	
39		1		1			2	4	1			2	1	5	
40	2	3		1			4	8	1		1	1			
41	782	396	1	7	2	7	272	79	148	26	37	47	18		6
42	30	94		6		2	106	35	212	13	27	40	424	1	
43	47	162		2	2	4	152	40	81	31	15	34	3		3
44	6	11					3	3	12	2	2	2	2		
51	1	58	5	8	3	2	8	64	10	18	4	12	83		3
52	8	524		19		6	87	104	5	19	761	23	54		1
53		109	1				3		7			10	12	136	2
54		3			1		7	6	467	4		36	1		
59		7				1			1	4	2	1	6		2
61		2					3	1	2	1			1		2
62									1						
71		1	3	4	2	6	4	25	21	2		7	4		33
72	1	11			1	27		11	3	10	1	7	4		88
73		1		1		2		2		4		2	2		10
74		5		2	1	23	8	3	4	3		2	5		71
75		10		3	1	1	2	8		2		2	2		12
81	1	20	2	2		6	6	2	2	143	1	10	2		26
82		7			1	2	1	1	6	22		4	1		16
83	1	21			1,817	2		3	5	70	3	4	11		3
91		17	1,008					9	4	3		2	5		1
92					3			2	1			1			2
93		59			6	3	1	1	4	23	1	1	1		5
94		279					3	3			9	2	26		
96	2	70	1		1		3	2	6	5		4	15		9

Auto-Classification



Auto-Classification – Input Text

df_test - DataFrame

Index	label	text
0	0	domestic helper household duties
1	0	account director. apac looks after client accounts
2	0	call center agent answering calls. taking feedback.
3	0	assistant professor teach history and religious studies in university..conduct research and publish books and journal articles.
4	0	senior lecturer teach music in an arts college.
5	0	manager for backend operations manager for backend operations
6	0	remisier stocks dealing
7	0	cto manages the product. engineering. qa and design functions of the company
8	0	head of advanced analytics manages the analytics team and supports business functions
9	0	secretary secretarial support
10	0	storeman managing and issuance of stocks
11	0	research associate doing research data entry
12	0	no job assignment maritime survey and consultancy
13	0	event co ordinator prepare menu. communicating and liaise with client and arrangements of event.
14	0	infant educator specializes in providing an educational environment in a preschool to babies and toddlers from the age of 3 months up to 3 years old
15	0	sales assistant screening and temperature taking. replenish products. receiving stock and ticketing.
16	0	business control manager control management for business including project work
17	0	managing director & global head. enterprise risk data & reporting management. direct onshore and offshore teams globally involved in data strategy. management. risk reporting and risk appetite.
18	0	senior legal counsel advise on legal matters in relation to employer
19	0	domestic helper household duties
20	0	partner in a consulting firm sells and supervises consulting engagements
21	0	supply chain manager manages supply chain issues at regional level
22	0	domestic helper household duties
23	0	domestic helper household duties
24	0	contract work marketing consultation work

Auto-Classification – Label Score

df_result - DataFrame

Index	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	4.71452e-05	2.33558e-05	6.29759e-05	0.000117798	1.90299e-05	0.000199289	6.47638e-05	1.26709e-05	0.000187032	6.95338e-05	2.47416e-05	3.51994e-05	5.24905e-05	0.00144157
1	9.35079e-06	0.000142809	1.83163e-05	1.19893e-05	4.51242e-05	4.73843e-05	9.84758e-06	4.68386e-05	4.00391e-05	6.06258e-05	0.00011624	0.00879323	3.41067e-05	0.00199632
2	0.000182913	0.000597833	0.000255389	0.000346321	0.00022486	0.000213637	0.000334011	0.000318948	0.000299808	0.000268735	0.000728896	0.00146165	0.00054465	0.00431029
3	0.000227597	0.000961932	0.000156284	0.000262297	0.000257759	8.43542e-05	7.26602e-05	0.000394457	0.00010896	0.000128842	0.000422265	0.000453312	0.000270382	0.000321035
4	0.000130336	0.00106851	9.50255e-05	0.000143923	0.000284899	6.63242e-05	5.73849e-05	0.000443308	8.44925e-05	0.000105453	0.000326043	0.000211365	0.000178264	0.000135506
5	1.69138e-05	0.000188781	1.62376e-05	2.96215e-05	9.85277e-05	8.77577e-05	9.82289e-06	4.86251e-05	0.000156326	0.000125136	0.000119804	0.00387404	6.28548e-05	0.0054249
6	7.04591e-05	0.000452981	6.92078e-05	4.68419e-05	0.000100646	0.000126536	4.11794e-05	0.000270922	7.98094e-05	0.0001404	0.000226324	0.00244266	0.000149023	0.000550661
7	1.82993e-05	0.000269316	1.50605e-05	2.80302e-05	0.000465532	0.000251435	1.85095e-05	0.000113691	0.000192269	0.000718929	0.000429395	0.00451068	4.88922e-05	0.0010593
8	4.15085e-05	0.00174324	4.80442e-05	2.87385e-05	0.000323993	0.000116998	2.29471e-05	0.000303116	0.00012861	0.000199349	0.000486488	0.0067266	0.000161614	0.000418147
9	4.65332e-05	0.00011993	2.48425e-05	5.88112e-05	4.143e-05	9.06498e-05	1.85571e-05	2.29926e-05	6.65633e-05	7.13941e-05	6.53387e-05	0.000609956	5.98608e-05	0.0214345
10	1.41559e-05	0.000107483	1.40441e-05	4.95149e-05	2.55372e-05	0.000143193	3.43595e-05	6.60071e-05	5.13604e-05	0.000114566	6.93065e-05	0.000743803	4.14375e-05	0.00342676
11	9.11608e-05	0.0157519	0.000135115	0.000138787	0.000955933	0.000130095	5.37452e-05	0.00108136	0.000193492	0.000492705	0.000210863	0.000749937	0.000281615	0.000538752
12	0.000555199	0.00309858	0.000324295	0.00117267	0.00522243	0.000472618	0.000648452	0.0111811	0.000469145	0.00190587	0.00058781	0.000590947	0.000916022	0.000130266
13	5.20834e-05	7.63519e-05	2.59284e-05	9.48036e-05	3.72055e-05	0.00013048	3.50103e-05	3.43066e-05	0.000102136	0.000165584	5.73307e-05	0.00071297	6.0634e-05	0.00305083
14	4.32792e-05	0.00013321	5.04143e-05	6.69517e-05	3.89999e-05	7.57362e-05	1.72461e-05	4.01276e-05	6.42641e-05	5.69684e-05	1.25698e-05	6.07614e-05	4.79169e-05	0.000264649
15	4.03836e-05	0.000280139	2.44369e-05	7.1059e-05	5.09788e-05	0.000175582	7.37686e-05	8.81712e-05	6.20943e-05	0.000258984	0.000182913	0.00208175	7.36436e-05	0.00243553
16	3.08995e-05	0.000169982	2.81453e-05	5.29046e-05	0.000149241	0.000180321	3.17821e-05	9.93738e-05	0.000106689	0.000274581	0.000546356	0.0206053	0.000138548	0.000859682
17	1.75676e-05	0.000159126	1.74964e-05	1.85313e-05	8.09625e-05	8.76085e-05	8.73262e-06	5.06969e-05	9.23482e-05	0.00013332	0.000110072	0.00469931	4.1482e-05	0.00188705
18	0.000218254	6.94396e-05	0.000116605	9.0094e-05	0.000126508	9.5947e-05	6.88687e-05	0.000106999	0.000108409	0.000133548	0.000134869	0.000149124	0.000101109	0.000515
19	4.71452e-05	2.33558e-05	6.29759e-05	0.000117798	1.90299e-05	0.000199289	6.47638e-05	1.26709e-05	0.000187032	6.95338e-05	2.47416e-05	3.51994e-05	5.24905e-05	0.00144157
20	1.95593e-05	0.000231665	1.88618e-05	2.27306e-05	0.00010472	0.000138985	1.76444e-05	6.04361e-05	7.08992e-05	0.000164456	0.000484536	0.00894644	6.68891e-05	0.00194872
21	2.0325e-05	0.000297022	1.96823e-05	4.69941e-05	0.000308528	0.00033979	4.9814e-05	0.000191063	0.000197367	0.000482905	0.000654684	0.00648193	0.000109669	0.00073169
22	4.71453e-05	2.33559e-05	6.29759e-05	0.000117798	1.903e-05	0.000199289	6.47638e-05	1.26709e-05	0.000187032	6.95338e-05	2.47417e-05	3.51995e-05	5.24906e-05	0.00144157
23	4.71453e-05	2.33559e-05	6.29759e-05	0.000117798	1.903e-05	0.000199289	6.47638e-05	1.26709e-05	0.000187032	6.95338e-05	2.47417e-05	3.51995e-05	5.24906e-05	0.00144157

Auto-Classification – Top 3 Labels

df_map_result - DataFrame

Index	label	text
0	[91309, 51501, 91160]	domestic helper household duties
1	[12111, 13460, 24139]	account director. apac looks after client accounts
2	[42230, 42245, 96291]	call center agent answering calls. taking feedback.
3	[23101, 23102, 23109]	assistant professor teach history and religious studies in university..conduct research and publish books and journal articles.
4	[23101, 23102, 23300]	senior lecturer teach music in an arts college.
5	[13304, 13301, 12112]	manager for backend operations manager for backend operations
6	[24141, 24341, 24143]	remisier stocks dealing
7	[13211, 13291, 25160]	cto manages the product. engineering. qa and design functions of the company
8	[25112, 24312, 24131]	head of advanced analytics manages the analytics team and supports business functions
9	[41201, 41103, 41101]	secretary secretarial support
10	[43212, 43211, 43219]	storeman managing and issuance of stocks
11	[21347, 24291, 31413]	research associate doing research data entry
12	[31157, 21494, 21652]	no job assignment maritime survey and consultancy
13	[33320, 96251, 42242]	event co ordinator prepare menu. communicating and liaise with client and arrangements of event.
14	[36100, 53120, 53113]	infant educator specializes in providing an educational environment in a preschool to babies and toddlers from the age of 3 months up to 3 years old
15	[52202, 43212, 52190]	sales assistant screening and temperature taking. replenish products. receiving stock and ticketing.
16	[12212, 11203, 33221]	business control manager control management for business including project work
17	[24132, 12111, 13460]	managing director & global head. enterprise risk data & reporting management. direct onshore and offshore teams globally involved in data strategy. management. risk reporting and risk appetite.
18	[26112, 26111, 26119]	senior legal counsel advise on legal matters in relation to employer
19	[91309, 51501, 91160]	domestic helper household duties
20	[24212, 24211, 12212]	partner in a consulting firm sells and supervises consulting engagements
21	[13241, 13242, 13211]	supply chain manager manages supply chain issues at regional level
22	[91309, 51501, 91160]	domestic helper household duties
23	[91309, 51501, 91160]	domestic helper household duties
24	[33229, 33223, 12211]	contract work marketing consultation work

Results

**Good
Predictions
(70%)**



- ✓ Intervention not required!
- ✓ Robust for certain categories



**Bad
Predictions
(30%)**

- ✓ Intervention required!
- ✓ Information may be insufficient to make a good categorical decision

Conclusion

With the implementation of the automatic classification model, it has led to a reduction in man-hours spent on data collection and brought about cost savings to the organization.





Thank You