Discovering Indonesian Digital Workers in Online Gig Economy Platforms

A. Labib Fardany Faisal, Yudho Giri Sucahyo, Yova Ruldeviyani, Arfive Gandhi
Faculty of Computer Science
Universitas Indonesia
Depok, Indonesia
a.labib61@ui.ac.id, yudho@cs.ui.ac.id, yova@cs.ui.ac.id, arfive.gandhi@ui.ac.id

Abstract— **Having a rapid growth across the world, Online Gig Economy (OGE) has the potential to reduce unemployment in Indonesia. It offers flexible working arrangement, flexible recruitment and new job types. Unfortunately, current existing economic and labor measurement systems are still not suitable to measure OGE distribution in Indonesia, especially for digital gig workers. The goal of this research is to portray Indonesian digital workers in OGE Platforms. This research relied on web crawling and web scraping for data collection combined with Automatic Text Classification (ATC) for data aggregation and classification. By delving nine platforms, 2,062 active gig workers were captured from 171,033 Indonesian users. Their profile was distributed into several dimensions: affiliated platforms, work fields, provinces, and paid salary. The result shows that most of gig workers were categorized in creative and multimedia. Considering IDR 3.4 million as the average of gig workers' paid salary, gig economy offer competitive and promising alternative for society to get money. These findings showed the significant role of internet to achieve better live and reduce unemployment.**

Keywords—digital worker; gig economy; gig worker; data classification; web crawling; web scraping; classification

I. INTRODUCTION

Internet growth has encouraged the appearance of gig economy as trend in global digital business. Many countries have identified its appearance and promoted it as new advantages to empower their citizens' ability. Abraham, et.al [1] identified rapid growth of gig workers in US which indicated by increasing tax-filing of self-employee. Online Labour Index (OLI) as established by Oxford University announced 37.50 percent of increasing growth of gig workers during August 2016 to November 2017 [2].

OLI's result showed United States, United Kingdom, India, Australia, and Canada have big domination on gig workers' profile [3]. It signalizes bigger chances for gig workers from those countries to get income without physical border limitation. They can use internet as 'door' to sell their ability in global virtual marketplace. Surprisingly, India is one of the top-five although it is a developing country with high rate of unemployment.

Indonesia as one of biggest countries should realize this opportunity to explore and exploit potential of gig economy. With fourth rank on populations below China, India, and USA, Indonesia has a chance to encourage its people to take advantage of gig economy. The digital divide in Indonesia is also getting better with 64.8% (171.17 million) people are internet users [4]. On the other side, Indonesia still faced 5.5% unemployment [5] [6]. Fluctuated macro-economy often affects low sustainability of big corporates. Therefore, gig economy as flexible work scheme offers promising opportunity to reduce unemployment rate and increase individual income to achieve better live.

Unfortunately, the Government of Indonesia still has not enough statistics about Indonesian gig workers. Without adequate statistics, government cannot formulate decent and proper strategic policies and technical regulations to control economy. This research contributes statistics about gig workers from reputable digital platforms. It also enables qualified and necessary program from government and universities to develop the capacity of gig workers. By doing so, Indonesian gig workers can be more competitive and contribute in individual wealthy.

The rest of this paper is organized using sections as follows. Section II delivers related literatures review. Section III reveals the methodology. Results, analysis, and implications are presented in Section IV. Finally, Section V and VI narrate the conclusions and recommendations.

II. LITERATURES REVIEW

A. Gig Economy Classifications

Simply put, gig economy is defined as work scheme to do project without fixed employment relationship. It relates with several similar terminologies, such as online labor, online outsourcing, and online labor [7] [8]. Refer to Schmidt classification in [8]; a rigid taxonomy has been constructed to decompose digital platform labor into web-based and location-based. Both of them are divided by amount of involved labor: individual and collective/crowd.

Heeks revealed more simple taxonomy about gig economy as shown in Fig. 1. It distinguishes gig economy into physical (PGE) and online gig economy (OGE) which medium and way to deliver products/services as key difference. As an example, Grab is classified as PGE since the gig worker serves client in physical environment directly. On the other side, Upwork facilitates gig workers to deliver their product through digital platform so that it is classified as OGE.

978-1-7281-1655-6/19/$31.00 ©2019 IEEE

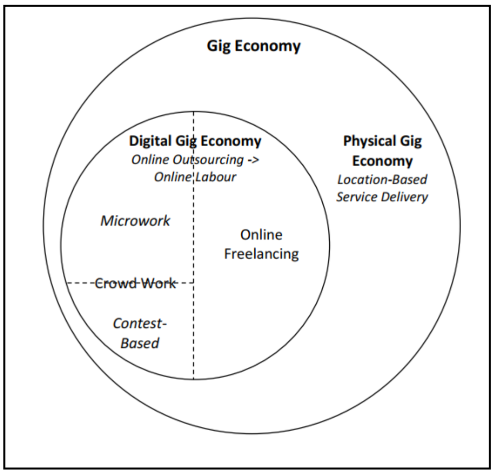

Fig. 1. Classification of Gig Economy [7]

B. Previous Online Gig Economy Measurement

Kassi and Lehdonvirta in [3] have introduced Online Labour Index (OLI). It measures OGE globally by aggregating five largest online labor platforms (Freelancer.com, Guru.com, MTurk.com, Peopleperhour.com and Upwork.com) using API and web scraping. Their initiative is published in real time at ilabour.oii.ox.ac.uk/online-labour-index/. It comprises these data dimensions; number of job openings, occupations, and national origin. Kassi in [2] has added data from the supply side (workers) with the dimensions of the work field and country of origin as well.

Meanwhile, Ross, Zaldivar, Lilly Irani, & Tomlinson in [9] deliver more complete data than OLI. It includes number of workers, country of origin, gender, income, time of system usage, age, education and others. This more complete data is generated by spreading online survey work to the MTurk.com with a predetermined wage. Unfortunately, the provided data is limited for MTurk.com only.

C. Web Crawling and Scraping

Generally, web crawling focuses on techniques to collect content of webpage from a root Uniform Resource Locator (URL) using web spider. Kadam et.al in [10] told that web crawling works by passing and collecting all webpages as nodes. It performs parallel and multi-thread to generate inter-thread synchronization and time efficiency. It can also distinguish visited node and unvisited node to anticipate duplication. Turland in [11] defined web scraping is data taking process in semi-structured document from internet. The data taking process can be manipulated by selecting the desired data based on business needs. To conduct this manipulation, the location of desired data in webpages should be identified. Steps to do web scraping can be cascaded through [12]: Create scraping template; Explore site navigation; Automate navigation and extraction; and Extracted data and package history.

In this research, web crawling and scraping become main technique to collect primary data from online digital platforms for online gig economy, such as Fiverr, Upwork, and Fastwork.

D. Automatic Text Classification (ATC)

ATC highlights categorization on digital documents into some defined class [13]. It employs classification algorithms, such as support vector machine, probabilistic methods, decision tree, [13] and vector space model [14]. Wijewickerma and Gamage in [13] stated text pre-processing, training set, and classification algorithm usage as steps on ATC. This research also mixes some extra technique to strengthen similarity analysis text on dataset: string distance, cosine similarity, and Term Frequency-Inverse Document Frequency (TF-IDF).

III. Methodology

A. General Method

This research had run quantitative approach by aggregating primary data from various websites. Web crawling and scraping were employed as main technique for data collection based on necessary information criteria. Information visualization followed the automatic classification to deliver insightful mapping. Fig. 2 illustrated this scenario. Since primary data were obtained through web crawling and web scraping without field experiment, this research was also categorized as survey and desk research.

B. Constructing the Information Criteria

Information criteria represented which variables are included or excluded from aggregation processes. It aimed to align different variables as contained in original sources. Without information criteria, aggregation processes can lose reliability since some information is not required. Therefore, aggregation processes perform more effectively.

Information criteria were cascaded from National Workforce Survey (Sakernas) as the baseline of workforce statistics in Indonesia. It comprised numbers of workers, unemployment, labor forces, and also population aged 15 years or more. The availability of Sakernas's dimension should be checked in gig economy platform to ensure alignment between necessary national standard with platforms' ability to provide information. Its result are shown in Table I.

In Sakernas, digital labor was not described explicitly, but can be interpreted from other terminologies. There were seven categories of employment as defined in Sakernas: self-employed, trying to be assisted by temporary workers, trying to be assisted by permanent workers, workers / employees, free workers in agriculture, workers free in non-agriculture and family / unpaid workers. Based on those categories, digital workers can be found as mixed model among self-employment, labor/employees, and free workers.

In Field of Work, this criterion was related with similar terminologies in any platforms, such as skills, specialties, or portfolios. Each platform also used various names for similar field of work. Therefore, this research standardized those using classifications from the Online Labor Index (OLI). For examples, small tasks which require basic human ability (such as data entry and image classification) was categorized as clerical and data entry category. Sales and marketing support were categorized as online advertising and marketing. Jobs in the professional services category usually required formal education and knowledge of the organizational system.

TABLE I. INFORMATION CRITERIA JUSTIFICATION

Criterion	*Argumentation*	*Y/N*
Number of workers	This research counts the digital labor with last project has been held no more than 28 days as mandatory requirement	Yes
Working hours	All platforms do not display data for this criterion	No
Field of work	Classification algorithms are applied with skill data sources, specializations and completed projects so that workers are mapped into one of the six occupational fields.	Yes
Level of education	Although some platforms set it as optional field, this criterion fulfills the sufficient samples	Yes
Job status	Based on definition of Sakernas model, digital worker is mixed status among self-employment, labor/employee, and free workers. Hence, all statistics about their existence is relevant.	Yes
Age	All platforms do not display data for this criterion	No
Province	All platforms (except Fiverr.com, Sribulancer.com and Fastwork.com) display districts / cities from workers so that provinces can be tracked.	Yes
Average wages per month	Actually, none of platform displays monthly wages explicitly. But, they have supporting data as an approach, such as total received money and working duration. Therefore, this criterion can be estimated.	Yes

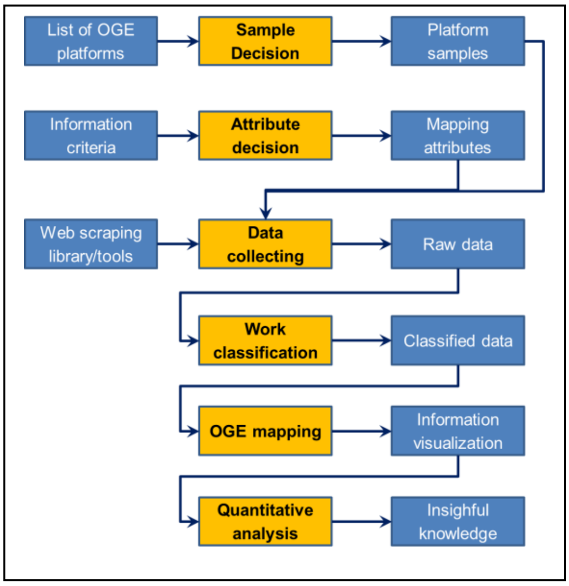

Fig. 2. Methods

C. Sample and Data Processing

This research focused to some online gig economy platform to achieve efficiency in aggregation processes. They were chosen by considering their traffic in Alexa, availability of shared information (includes gig workers' profile), experience in empirical experiment in previous researches, and variety of offered works. Second requirement affects Mturk.com, Clickworker.com, and Crowdflower.com were eliminated. This research also excluded contest-based application (such as 99Designs) due to different digital business model. Finally, nine platforms were selected for this research's experiment Upwork.com (GP.01), Freelancer.com (GP.02), Fiverr.com (GP.03), Peopleperhour.com (GP.04), Guru.com (GP.05), Sribulancer.com (GP.06), Projects.co.id (GP.07), Fastwork.id (GP.08), and Microworkers.com (GP.09).

IV. RESULTS, ANALYSIS, AND IMPLICATIONS

A. Data Collection Processes

In this process, several iterations need to be done to get stable hyperlinks on each platform for web crawling and scraping processes. The founded hyperlinks are listed in Table II. Each platform will be given two scripts. The first script works to crawl worker hyperlinks and store them as strings in .txt files (crawling process). The second one works to scrap/collect worker profiles based on hyperlinks in .txt file produced by first script (scraping process).

Scripts are written on Javascript by using Puppeteer library so they have ability to control an internet browser autonomously as automatic agent. The library has some APIs to open a website by custom URL, clicking buttons or hyperlinks, writing forms, scrolling pages and applying settings of the browser (e.g. proxy address and window size). A script collects data like a human does, as it must go to every pages, click some buttons to open more detailed information, copy the information text needed and store them in text files.

As an example of web crawling and scraping process, in Fastwork, webpages for worker searching are structured as parent and child URL. Parent URL refers to default hyperlink to access Fastwork while child URL is associated with any sub-folder hyperlink, such as /categories/graphics-design. In this research, an automatic agent is implemented to access entire sub-folders by tracking all available child URL in homepage using Inspect Elements feature as provided by internet browser. Tag selector should be identified and adopted in the automatic agent to optimize web scraping. This scenario is illustrated in Fig. 3.

TABLE II. WEBPAGE LINKS FOR WEB CRAWLING AND SCRAPING

Plat-form	Webpage for User Searching (Web Crawling Process)	Webpage for User Profile (Web Scraping Process)
PG.01	/o/profiles/browse/?loc=indonesia/page=[n]	/o/profiles/users/[USER_SPECIFIED]
PG.02	/freelancers/indonesia/all	/u/[USER_ SPECIFIED]
PG.03	/d/freelancers/l/indonesia/pg/[n]	/freelancers/[USER_ SPECIFIED]
PG.04	/hirefreelancers?location=ID&page=[n]	/freelancer/[USER_ SPECIFIED]
PG.05	/id/bf/freelancer/v4?page=[n]	/id/bf/freelancer/v4
PG.06	/public/browse_users/listing	/public/browse_users/view/[USER_ SPECIFIED]
PG.07	/categories/[JOB_CATEGORY]?source=hp_subcat_sec&ref=seller_location%3AID&filter=rating&offset=0 &page=[n]	/[USER_SPECIFIED]
PG.08	/[JOB_CATEGORY]?page=[n]	/user/[USER_SPECIFIED]
PG.09	/hm_pool.php?Sort=TASKS&Id_badge=[JOB_CATEGORY]?&Countrycode=id	/userinfo.php?Id=[USER_ SPECIFIED]

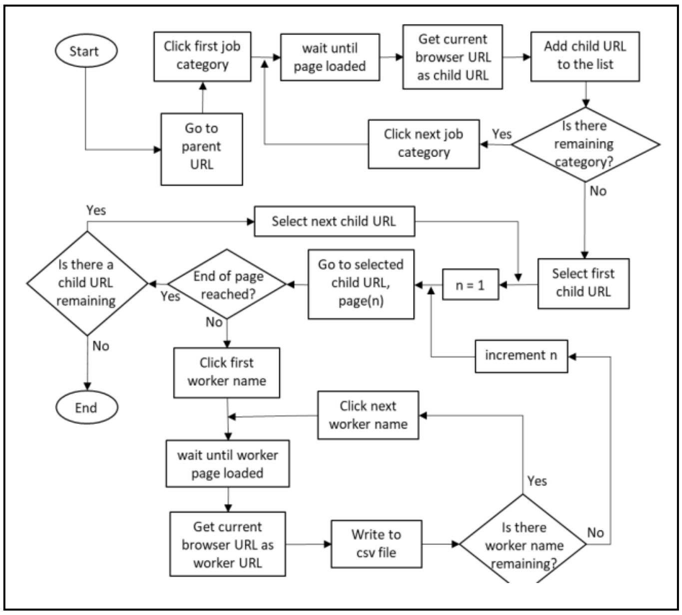

Fig. 3. Web Crawling Scheme on PG.08

During the scraping process, automatic agent visits to each worker's profile page based on URL in the .txt file. Same as former process, automatic agent collects worker's data based on identified tag selector in the HTML document. Some data can be obtained directly (such as worker name), while others require a button click then the data appears (such as project names). When raw data has been successfully generated, it was stored in a standard table. It comprises these attributes: ID, platforms, worker name, city, skill, specialization, project names, latest project date, latest education, and monthly salary. Some attributes are blank since the platform has no related raw data while some attributes cannot be filled automatically. The second case is occurred since the applications hide them. Therefore, advance searching should be demonstrated to anticipate it. Data with this handicap situation are labeled as implicit data while the captured data without problems are labeled as explicit data.

This research discovers many implicit data, especially related with monthly salary. All platforms do not display it clearly. They provide raw data about paid salary on each worker's performed project. This research requires more complex formula to convert the paid salary into monthly salary by reviewing the work scheme in each platform. As simplest example, annual paid salary as exhibit in PG.03 can be converted into monthly paid salary by dividing it with 12 months. Some platforms, such as GP.01 and GP.04, mandate more rigid formula due to tracing project period.

B. *Classification of Work Field*

Work Field cannot be shown in all webpages for worker searching. Moreover, available Work Fields have different naming among platforms. Hence, this research needs classification to standardize captured data about Work Field by considering specializations, skills, and project names variables. Since the data volume is too large, this research uses Automatic Text Classifier (ATC) by mixing Jaro-Winkler distance technique, TF-IDF and cosine similarity. Basically, this research actualizes Work Field classification as introduced by [3] in OLI. It comprises six classes which include many project types as follows:

- Professional services (CL.01): accounting, consulting, financial planning, legal services, human resources, and project management;
- Clerical and data entry (CL.02): customer services, data entry, transcription, tech support, web research, and virtual assistant;
- Creative and multimedia (CL.03): animation, architecture, audio, logo design, photography, presentations, video production, and voice acting;
- Sales and marketing support (CL.04): ad posting, lead generation, SEO, and telemarketing;
- Software development and technology (CL.05): data science, game development, mobile development, QA and testing, server maintenance, software development, web development, and web scraping;
- Writing and translation (CL.06): academic writing, article writing, copywriting, creative writing, technical writing, and translation.

ATC has discovered hundreds skills from gig worker profiles and categorized them into related project type. Table III exhibits sample of tabulation for mapping related skill into classes and project types. Since PG-08 used Indonesian language, their project types are still labeled in Indonesian.

TABLE III. MAPPING OF SKILL INTO CLASSES AND PROJECT TYPES

Class	Project Type	Numbers of Related Skill	Plat-form
CL.05	Websites it and software	494	PG-02
	Engineering and science	84	PG-02
	Mobile phones and computing	22	PG-02
	Pemrograman web	7	PG-08

C. *Performance Analysis*

This research activates some metrics to measure performance quantitatively. First metrics focuses of how match classification scheme works on gig worker profiles. 433 samples are brought by considering proportional among classes. Comparison between their classification using ATC and manual classification is tabulated as Confusion Matrix as shown in Table IV. Its row and column represent ATC-based and manual-based respectively. Numeric data in cell *m-row x n-column* describes how many gig workers are classified as m using ATC-based and as n using manual-based. Shaded cell indicates correct matching between same classes. Using Confusion Matrix formula, this research can produce detail variable for other metrics: accuracy, precision, recall, and F1 score. Produced value for these variables are presented in Table V. General accuracy can be measured by count all values in shaded cell with total sample. It results 83.8% which is better value than OLI system with 72%.

TABLE IV. CONFUSION MATRIX

	CL.01	CL.02	CL.03	CL.04	CL.05	CL.06
CL.01	177	5	2	6	0	0
CL.02	3	69	6	3	1	0
CL.03	3	0	47	0	3	2
CL.04	7	1	1	36	1	0
CL.05	3	0	1	5	27	1
CL.06	10	1	3	1	1	7

TABLE V. PERFORMANCE METRICS

Class	Accuracy	Precision	Recall	F1 Score
CL.01	0.91	0.93	087	0.90
CL.02	0.95	0.84	0.91	0.87
CL.03	0.95	0.85	078	0.82
CL.04	0.94	0.78	0.71	0.74
CL.05	0.96	0.73	0.82	0.77
CL.06	0.96	0.30	0.70	0.42

D. Distribution Analysis

This research has collected 171,033 Indonesian users from nine platforms. This data is sample only due to various barriers in some platforms, such as incomplete access and incomplete profile. Since this research defines active gig workers should be involved in project no more than 28 days from web crawling and scraping, 2,062 users only is categorized as active Indonesian gig workers (see Table VI). Most of Indonesian gig workers are active in PG.03, PG.01, and PG.07, namely Fiverr, Upwork, and Projects.co.id. Significant numbers and percentage on PG.03 indicates Fiverr as the most active and potential gig platform for Indonesian to hunt projects.

This research has also distributed the raw data into some dimensions based on work field, provinces, and paid salary. Presented in Fig. 4, most of Indonesian gig workers has specialty in creative and multimedia (49.47%). By specifying into platforms as shown in Fig. 5, similar domination is shown in five out nine platforms state similar domination. It signalizes high potential chance for Indonesian people to strengthen creative industry as global trending. Different distribution is occurred in Miroworkers.com which dominated by Clerical and Data Entry. It was understandable since it focuses on simple job with large transaction.

As depicted in Fig. 6, this research groups Indonesian gig workers into provinces. It results Jawa Barat, Jawa Timur, and DKI Jakarta as top-three with more than 15% respectively. Moreover, all top-six provinces are in Java Island. It portrays large a gap between Java societies with others. Therefore, potential market of gig economy needs huge promotion to encourage participation from other provinces.

Paid salary as the last classification is revealed in Fig. 7 (each category represents million in Indonesian Rupiah). Almost 50% gig workers have been paid 1 million or less. The higher paid salary is received by the fewer gig workers. This domination may be affected by gig workers' participation in clerical and data entry category which is paid with low salary. Interestingly, there are gig workers can grab more than 5 million which means more than any minimum regional salary in Indonesia. It proofs that gig economy offer chance get promising income to achieve better live. Moreover, statistics notice gig workers can get more than 19 million. Overall, 3.4 million was taken as average paid salary.

TABLE VI. GENERAL STATISTICS OF INDONESIAN USERS

Class	Indonesian Users	Active Indonesian Gig Worker	Percentage (%)
PG.03	837	557	66.55
PG.01	13,945	449	3.22
PG.07	115,558	428	0.37
PG.02	16,639	187	1.12
PG.06	9,841	181	1.84
PG.09	1,017	176	17.31
PG.04	464	40	8.62
PG.08	1,456	38	2.61
PG.05	11,276	6	0.05

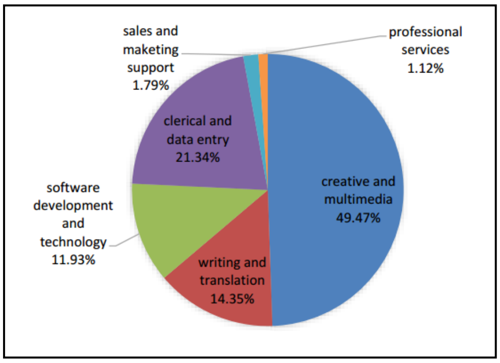

Fig. 4. Distribution of Indonesian Gig Worker's Work Fields

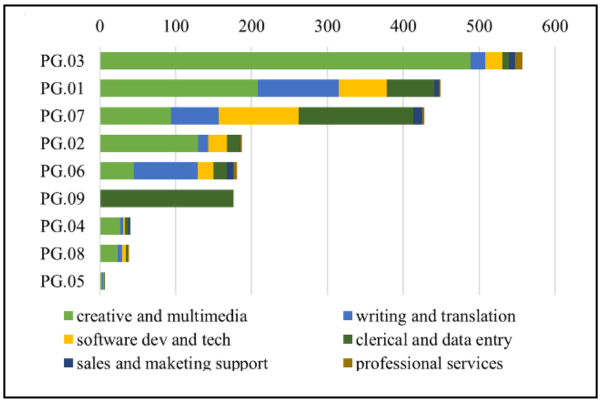

Fig. 5. Indonesian Gig Worker's Work Fields based on Platforms

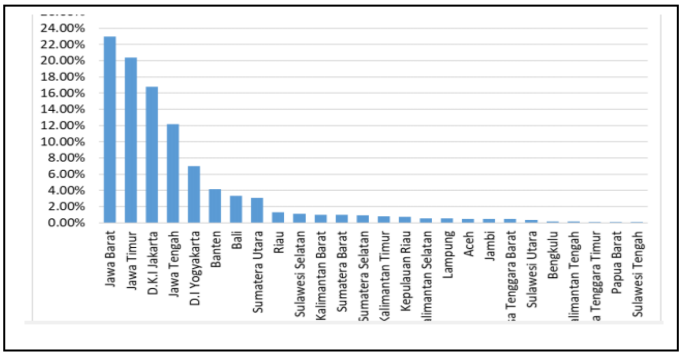

Fig. 6. Distribution of Indonesian Gig Worker's based on Provinces

V. CONCLUSIONS

This research has successfully discovered profile of Indonesia gig workers as digital worker who actively do temporary projects using online gig economy platforms. This contribution was derived by synthesizing web crawling and web scraping for data collection processes with ATC for data classification. It portrayed current state of Indonesian gig workers as novelty so that significance of its growth can be monitored more effectively in the future.

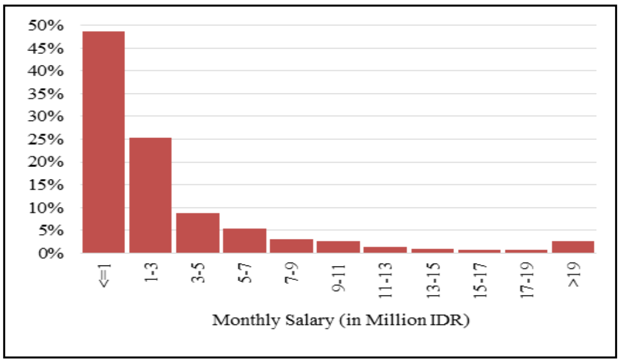

Fig. 7. Distribution of Indonesian Gig Worker's based on Paid Salary

Through nine sample platforms, 2,062 active gig workers were identified out of 171,033 Indonesian users. With 83.3% general accuracy using Confusion Matrix, this research found most of them are in creative and multimedia category (49.47%). Fiverr, Upwork, and Projects.co.id are the favorite platforms since their high gig workers' participation. It also results two patterns as big challenges to improve gig economy ecosystem in Indonesia: low participation of gig workers from non-Java Island and paid salary is dominated by 1 million or less. 3.4 million as average paid salary indicated gig economy can offer competitive alternative for people to get money other than work as fixed-employment.

VI. RECOMMENDATIONS

Based on performed experiment, this research suggests several recommendations. First, aggregation and classification processes should be enhanced by optimizing the algorithms. It enables more-frequent monitoring so that growth of gig workers can be evaluated more effectively. Second, ABG (Academician-Business-Government) should follow up the insights about gig economy growth in Indonesia depending on their roles. Academician need to adjust current curriculum with the skills as required by gig worker since gig economy has become promising work model. Moreover, government should formulate decent policies to encourage gig workers' ability. Third, this research also promotes future outlook to expand the usage of algorithm into physical gig economy, such as ride-sharing, food delivery, mentoring, mechanics, etc.

ACKNOWLEDGMENT

This research was supported by Universitas Indonesia under the PIT 9 Grant 2019 of "Optimization of Digital Business Contribution through Platform-based Project Management Model" (No: NKB-0014/UN2.R3.1/HKP.05.00/ 2019). We would like to express our gratitude to the Faculty of Computer Science and the Directorate of Research and Community Engagement, Universitas Indonesia.

REFERENCES

[1] K.G. Abraham, K. Sandusky, J.C. Haltiwanger, J.R. Spletzer. "Measuring the Gig Economy: Current knowledge and open issues", Tech. rept. 2017.

[2] O. Kassi, Measuring the supply of digital labour: How the OLI worker supplement is constructed, available on ilabour.oii.ox.ac.uk/measuring-the-supply-of-digital-labour-how-the-oli-workersupplement-is-constructed/

[3] O. Kassi, V. Lehdonvirta, "Online Labour Index: Measuring the Online Gig Economy for Policy and Research", Technological Forecasting & Social Change, pp. 1-8, 2018.

[4] Association of Indonesian Internet Service Providers, Penetration and Behavioral Profile of Indonesian Internet Users in 2018, 2019.

[5] Trading Economics, Unemployment Rate | Asia, available on tradingeconomics.com/countrylist/unemployment-rate?continent=asia

[6] BPS, Work Force Situation in Indonesia August 2017, BPS, 2017.

[7] R. Heeks, Digital Economy and Digital Labour Terminology: Making Sense of the "Gig Economy", "Online Labour", "Crowd Work", "Microwork", "Platform Labour", Etc. CDIGDI Institute, 2017.

[8] Schmidt, F. A. Digital Labour Markets in the Platform Economy. Bonn: FriedrichEbert-Stiftung, 2017.

[9] J. Ross, A. Zaldivar, LillyIrani, B. Tomlinson, "Who are the crowdworkers? Shifting demographics in Amazon Mechanical Turk", ACM CHI Conf. pp. 2863-2872, 2010.

[10] S. Kadam, S. Shinde, A. Sharma, S. Mali, "Price comparison of computer parts using web scraping", IJESC, pp. 16413-16415, 2015.

[11] M. Turland, Phparchitect's Guide to Web Scraping with PHP. Toronto: Marco Tabini, 2010.

[12] A. Josi, L.A. Abdillah, Suryayusra, "Penerapan Teknik Web Scraping pada Mesin Pencari Artikel Ilmiah", JSI Vol. 5, 2014.

[13] C.M. Wijewickerma,R. Gamage, "An ontology based fully automatic document classification system using an existing Semi-Automatic System", IFLA WLIC, pp. 1-13, 2013.

[14] A. Jain, N. Chauhan, A. Jain, V. Singh, N. Thakur, "Information retrieval using Cosine and Jaccard Similarity Measures in Vector Space Model", Int. J. of Comp. Applications, pp. 28-30, 2017.