

Eksplorasi Teknologi Pendukung Data Science untuk Industri Unggulan di Wilayah Kota Surabaya dan Sekitarnya

Rinabi Tanamal¹, Yosua Setyawan Soekamto², dan Trianggoro Wiradinata³

^{1,2,3}Program Studi Sistem Informasi, Universitas Ciputra, UC Town Citraland CBD, Surabaya, 60119
e-mail: ¹r.tanamal@ciputra.ac.id, ²yosua.soekamto@ciputra.ac.id, ³twiradinata@ciputra.ac.id

Submitted Date: March 26th, 2021

Revised Date: July 28th, 2021

Reviewed Date: May 31st, 2021

Accepted Date: August 08th, 2021

Abstract

Data science competence is still quite uncommon for most people, even though this competence is very important in managing business development strategies in various leading industrial fields. Through data science, companies can find patterns, perform clustering, calculate predictions more accurately to make decisions that have an impact on company excellence. Lack of knowledge about data science, researchers seek firstly in the industry to explore tools that are mostly used, convenience, challenges, measuring success tools and advantages. In this study, the researchers conducted a Focus Group Discussion by interviewing resource persons who worked as Data Scientists. Interview method to discuss aspects of Demographics, aspects of Technology-Task Fitness, and aspects of the Technology Acceptance Model. The results of this study conclude that the tools that are the mainstay of the industry in Surabaya and its surroundings area are Jupyter Notebook, Python, and Tableau. The success in using these tools still depends on the user's abilities, the results of user simulations, team discussions, and the evaluation process.

Keywords: data science; pattern recognition; clustering; prediction; decision making

Abstrak

Kompetensi *data science* masih cukup awam bagi kebanyakan masyarakat, padahal kompetensi ini sangat penting dalam mengatur strategi pengembangan bisnis di berbagai bidang industri unggulan. Melalui *data science*, perusahaan dapat menemukan pola, melakukan clusterisasi, menghitung prediksi dengan lebih akurat guna mengambil keputusan yang berdampak pada keunggulan perusahaan. Kekurangan pengetahuan mengenai *data science* ini, peneliti mencari tahu dahulu pada Industri untuk menggali tools yang mayoritas digunakan, kemudahan, tantangan, alat ukur keberhasilan dan keunggulan. Pada penelitian ini peneliti melakukan Fokus Group Discussion dengan menginterview para narasumber yang berprofesi sebagai Data Scientist. Metode wawancara untuk membahas aspek Demografi, aspek *Technology-Task Fitness*, dan aspek *Technology Acceptance Model*. Hasil penelitian ini menyimpulkan bahwa tools yang menjadi andalan industri di Surabaya dan sekitarnya adalah Jupyter Notebook, Python, dan Tableau. Keberhasilan dalam penggunaan tools tersebut tetap bergantung pada kemampuan pengguna, hasil simulasi user, diskusi tim, dan proses evaluasi.

Kata kunci: *data science; pengenalan pola; klusterisasi; prediksi; pengambilan keputusan*

1 Pendahuluan

Pada umumnya strategi bisnis dibuat dengan pengalaman pemilik usaha atau manajerial, tetapi seiring perkembangan zaman terjadi perkembangan strategi yang seiring juga dengan perkembangan teknologi. Sektor industri sudah harus mulai memanfaatkan *data science* dalam ERP mereka, sehingga meningkatkan efisiensi bisnis dengan lebih maksimal. Saat ini misalnya,

mulai banyak industri UMKM yang memanfaatkan teknologi perdagangan online (Tokopedia, Bukalapak, Blibli, JD.id dan lainnya) untuk memasarkan produk-produknya. Implikasi revolusi industri 4.0 dan berkembangnya *life style* industri menuntut dikembangkannya program studi yang berfokus ke *data science* yang mencetak *data scientist* yang ahli di era *digital economy* serta *e-commerce*.

Secara umum dapat dipahami bahwa *data science* merupakan irisan dari ilmu bisnis, statistik dan teknologi (*computer science*). Adopsi sistem informasi yang menggunakan *data science* menjadi peranan penting untuk mendukung percepatan penguasaan teknologi pendukung *data science* untuk industri. Adopsi sistem informasi tersebut juga harus didampingi dengan landasan teori, konseptual, dan empiris sehingga dapat memperkuat penguasaan teknologi pendukung *data science* tersebut. Belum adanya materi yang baku mengenai *data science* merupakan pemicu adanya studi ini yang sangat *urgent* untuk dilaksanakan.

Data Science adalah bidang interdisipliner yang mempelajari cara mengumpulkan, menyusun, dan mengubah data menjadi informasi dan pengetahuan untuk mendukung pengambilan keputusan berbasis data. Dalam konteks bisnis, *data science* berhubungan erat dengan *business intelligence* dan *data analytics* (Berman et al., 2018; Medeiros et al., 2020).

Adapun tujuan dari penelitian ini adalah untuk melakukan eksplorasi faktor-faktor yang memengaruhi percepatan penguasaan teknologi *data science* oleh Industri Unggulan di Surabaya dan sekitarnya.

Penelitian ini akan difokuskan untuk menunjukkan peranan teknologi informasi khususnya ERP untuk meningkatkan kompetitif dari industri, terutama dalam penelitian ini adalah industri unggulan di Surabaya dan sekitarnya.

Penelitian dilakukan di wilayah kota Surabaya dan sekitarnya agar dapat mempercepat pertumbuhan wirausaha dan agar Industri mampu menghadapi Era Industri 4.0. Pendidikan kewirausahaan merupakan instrumen yang digunakan untuk meningkatkan aktivitas kewirausahaan. (Bischoff, Volkmann, & Audretsch, 2018). Kewirausahaan juga dianggap sebagai sumber penting dari penciptaan lapangan kerja, pengurangan kemiskinan, inovasi dan social pembangunan serta daya saing ekonomi (Liñán, Rodriguez-Cohard & Rueda-Cantuche, 2011; Wu, Kuo, & Shen, 2013). Oleh karena itu, pertumbuhan yang stabil dari penciptaan bisnis diperlukan untuk kesejahteraan sosial serta untuk pembangunan ekonomi. (Ahmed, T., Chandran, V.G.R., Klobas, J.E., Liñán, F. and Kokkalis, P., 2020)

Dalam Jurnal "Factors affecting success of training companies" menjelaskan dalam artikelnya faktor apa saja agar training ke perusahaan dapat dikatakan sukses. Sebanyak 20 indikator mengindikasikan bahwa *Training* yang sukses juga

membutuhkan *Trainer* yang berkualifikasi dan memberikan kurikulum yang unik serta disesuaikan dengan kebutuhan *customer* di masing-masing industri. (Piotr, Batko, dan Wawak, 2017)

Sebagian besar dari para pendiri Google, Microsoft, Facebook, dan perusahaan-perusahaan sejenis adalah mahasiswa teknik yang mengalami perubahan dari disiplin ilmu teknik ke pola pikir kewirausahaan. Pendidikan kewirausahaan menawarkan kesempatan kepada masyarakat umum untuk berulang kali terlibat dalam suatu tugas, baik belajar tentang tugas tersebut dan bagaimana melakukannya, dan mengembangkan kepercayaan pada kemampuan mereka untuk melakukan tugas terkait dengan sukses di masa depan. Mereka juga terlatih untuk membangkitkan emosi positif dan memicu peristiwa yang mengubah pola pikir pembelajar, sehingga menghasilkan sikap positif terhadap kewirausahaan (Nabi et al., 2018). Oleh karena itu, diharapkan juga dengan melatih *data science* ke wirausahawan secara perlahan para akademisi juga akan menjadi pengusaha.

Salah satu faktor utama dalam meningkatkan jumlah inovator berbasis teknologi adalah sistem pendidikan yang dirancang dengan cermat berdasarkan faktor dominan yang mempengaruhi kesuksesan seseorang untuk menjadi wirausaha berbasis teknologi (Wiradinata dan Antonio, 2018). Studi ini secara singkat menggambarkan dan mengevaluasi metode yang diterapkan oleh departemen Teknik Informatika di Universitas Ciputra sejak 2013. Keterkaitan studi ini adalah aspek kekuatan dan kelemahan dari kurikulum yang saat ini telah ada dalam sistem pendidikan atau pelatihan dapat menjadi landasan dalam mempertimbangkan percepatan adopsi *data science* oleh industri unggulan di Surabaya dan sekitarnya.

2 Metodologi Penelitian

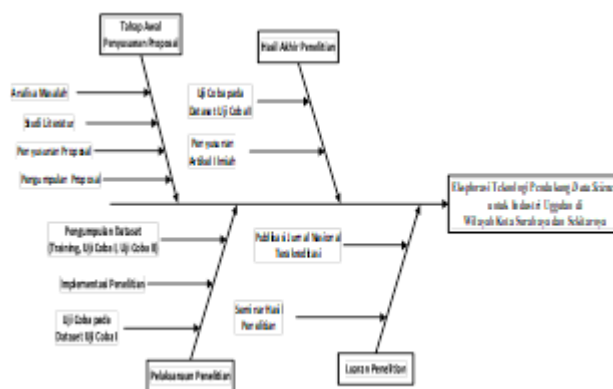
2.1 Metode Pengumpulan Data

Penelitian dilakukan dengan pendekatan kualitatif melalui studi literatur dan eksplorasi terkait faktor-faktor yang mempengaruhi percepatan penguasaan teknologi *data science* pada industri unggulan di Surabaya dan sekitarnya. Eksplorasi diterapkan dengan FGD dan atau mengundang Narasumber untuk memberikan informasi dan masukan untuk membuat draft sistem *data science* yang tepat guna untuk diterapkan pada industri unggulan di Surabaya dan sekitarnya. FGD adalah teknik di mana seorang peneliti mengumpulkan sekelompok individu

untuk membahas topik tertentu, yang bertujuan untuk menarik dari kompleks pengalaman pribadi, keyakinan, persepsi, dan sikap peserta melalui interaksi yang dimoderasi. (O. Nyumba, T., et al, 2018) Hasil yang diperoleh yaitu berupa draft yang sudah divalidasi oleh pakar sistem informasi *data science* yang tepat guna untuk industri unggulan di Surabaya dan sekitarnya.

2.2 Tahapan Penelitian

Penelitian ini merupakan penelitian tahap awal yang merupakan bagian dari analisa kebutuhan dalam *software engineering*. Oleh karena itu, tahapan penelitian ini dimulai dari eksplorasi teknologi pendukung *data science* seperti pengumpulan data dari survei atau *group discussion*, kemudian menyusun dan mengolah data tersebut untuk menentukan kesimpulan dan hasil.



Gambar 1. Detail Tahapan Penelitian dari Road Map Penelitian

Narasumber yang terpilih untuk mengikuti *group discussion* adalah sebagai berikut:

Tabel 1. Profil Narasumber

No	Nama	Pekerjaan	Perusahaan	Kode
1.	Alfan Alfian	Data Scientist	PT. SML Technology	AA
2.	Yoga P. Aliarham	Machine learning Engineer	Tiket.com	YP
3.	Theodore Adi Putra	Data Analyst	Alterra	TA
4.	Doni Rubriagatra	Senior Software Engineer	Zero One Group	DR
5.	Alto Salim	Enterprise Digital Technology Service	-	AS

6.	Lili Ayu Wulandh ari	Principle Data Scientist	Tiket.com	LW
7.	Farida	Chief Research Officer	RMDS Narasio Data	F

2.3 Uji Instrumen

Uji Reliabilitas: Peneliti melakukan uji reliabilitas dengan menggunakan *video recorder* pada aplikasi Zoom dan menggunakan *screenshot* untuk melakukan foto bersama saat proses FGD.

2.4 Metode Analisis Data

Peneliti menganalisis data dengan tiga tahap, yaitu:

a. Reduksi data

Tahap reduksi data diawali dengan melakukan *group discussion* dengan Narasumber yang berasal dari 7 industri unggulan di Surabaya dan sekitarnya untuk mendapatkan informasi yang diperlukan dalam penelitian. Dalam penelitian ini Narasumber memberikan informasi yang beragam dari pertanyaan yang diajukan oleh Peneliti. Selama proses *group discussion* berlangsung, Peneliti akan merekam jalannya diskusi dan menuliskan dalam bentuk *script*. Informasi yang didapatkan dari *group discussion* tersebut diseleksi untuk mendapatkan informasi yang diperlukan dalam penelitian (reduksi). Pada akhir penelitian, Peneliti akan mendengarkan ulang video rekaman untuk mengonfirmasi hasil reduksi.

b. Penyajian data

Penyajian data terdiri dari tiga aspek, yaitu aspek Demografi, aspek *Technology-Task Fitness*, dan aspek *Technology Acceptance Model*. Dengan menggunakan aspek *Technology Acceptance Model*, Peneliti dapat menemukan ketertarikan dan kebiasaan pengguna dalam menggunakan teknologi terkait. (Tanamal, R. (2019)

c. Penarikan Kesimpulan

Peneliti menarik kesimpulan dari masing-masing pertanyaan yang di berikan ke tujuh Narasumber yang akan di tunjukkan berupa table triangulasi. Peneliti kemudian merangkum kesimpulan dari hasil reduksi tujuh Narasumber dari beberapa industri unggulan di Surabaya dan sekitarnya.

2.5 Indikator/Pedoman Diskusi

Indikator yang menjadi acuan dalam penelitian ini adalah Demografi, *Technology-Task Fitness*, dan *Technology Acceptance Model*.

- Indikator Demografi terdiri atas:
- D-1: *Task* yang dikerjakan.
 - D-2: *Tools* yang digunakan.

Sedangkan indikator *Technology-TaskFitness* terdiri atas:

- TTF-1: Kesulitan dan tantangan.
- TTF-2: Alat ukur keberhasilan.
- TTF-3: Alasan utama pemilihan dan keunggulan.

Terakhir, Indikator *Technology Acceptance Model* terdiri atas:

- TAM-1: Tingkat kemudahan untuk dipelajari mandiri.
- TAM-2: Tingkat kegunaan untuk menyelesaikan suatu *task*.

3 Hasil dan Pembahasan

Tabel triangulasi disajikan berdasarkan masing-masing aspek sebagai berikut:

Tabel 2 Tabel Triangulasi *Demografi*

AA	YP	TA	DR	AS	LW	F	Kesimpulan
“Kalau saya lebih fokus ke riset nya jadi untuk mencari algoritma yang kira kira cocok dengan kasus yang dihadapi jadi lebih ke arah <i>research</i> sih” (AA-3)	“... intinya kita dari tim <i>machine learning</i> ini kita yang <i>develop</i> supaya model yang sudah di generate sama tim DS itu bisa dipakai user. Kita juga bikin <i>machine learning</i> pipeline supaya model itu tadi bisa <i>continous learning</i> . Supaya proses <i>learning</i> dari suatu model itu otomatis.” (YP-6)	“.... Ngerjainnya itu semua project yang berkaitan sama data. Artinya data ini bisa diolah jadi apa aja sih. Terus dibikin pengambilan - pengambilan keputusan gitu pak.” (TA-1)	“... kerjanya kurang lebih kayak <i>engineering manager</i> soalnya aku kayak end to end, jadi dari <i>software</i> itu dibuat di awal sampai <i>deployment</i> itu aku yang memanage. Jadi bener – bener kayak <i>engineering manager</i> , Cuma kalau ngomong konteksnya adalah <i>data science</i> . Di departemen yang aku jalani itu adalah bagian research and optimization. ” (DR-2)	“Nah, itulah kita membuatkan teknologi lagi, kedepannya supaya bisa masuk ke data, sehingga akhirnya datanya bisa dianalisa dengan benar. Jadi, secara garis besar EDTS itu bergerak di bidang <i>digital</i> dan didalamnya ada pembuatan produk, maupun data, gitu.” (AS-1)	“Di dalam <i>data science,</i> biasanya kita mulai dulu dengan <i>research,</i> gitu ya. Research, dalamnya kita akan mulai eksploratori data analisisnya terlebih dahulu, kemudian cari beberapa kemungkinan model yang kita bisa gunakan, kemudian kita lakukan eksperimen, kita compare hasil eksperimen nya tersebut, ..., baru nanti kita diskusikan lagi, nih ke <i>user</i> Kalau misalnya sudah oke, biasanya itu kita akan serahkan ke namanya <i>machine learning engineer.</i> ” (LW-2)	“Nah, disitu kami mulai mengerjakan yang pertama kali adalah mapping data-nya dulu, ... Setelah itu, kita melakukan standardisasi dan juga pembuatan metadatanya ... kemudian nanti baru kita analisis. ... data yang dimiliki tadi itu, tentunya sudah melalui tahapan pre-processing - atau cleaning data baru kita bisa menganalisis atau dibuatkan dalam bentuk visual supaya lebih mudah dipahami untuk <i>top management.</i> Jadi, fungsinya nanti memang untuk pengambilan kebijakan. ” (F-2)	Task yang dikerjakan di setiap industri berbeda-beda sesuai dengan Job descnya. Mulai dari <i>research</i> untuk mendapatkan algoritma yang tepat, <i>develop</i> model supaya bisa dipakai user, membuat <i>machine learning,</i> mengolah data sebagai dasar pengambilan keputusan, membuat teknologi lagi supaya data bisa dianalisis, melakukan eksperimen sebelum masuk ke <i>machine learning,</i> <i>mapping</i> data, melakukan <i>standardisasi,</i> membuat <i>metadata,</i> <i>cleaning</i> data, dan membuat visualisasi data.

<p>“Kalau untuk ke arah <i>modelling</i> sampai <i>production</i> itu kita menggunakan <i>Python</i> Jadi ya kalo hanya sekedar <i>exploratory data analysis</i> ...itu saya menggunakan <i>R</i>.... Nah <i>library</i> yang pakai itu namanya <i>Time Series Cluster</i> atau <i>TSClust</i>,” (AA-10) “... <i>Spyder</i> itu kayak <i>R studio</i> hampir mirip cuma nanti kalau kita buka <i>R notebook</i> berarti di <i>Spyder</i> nya ada kayak <i>Jupyter notebook</i>” (AA-25)</p>	<p>“.... Secara umum <i>machine learning</i> itu ... biasanya kita pakai <i>Python</i>, jadi kita bahasa pemrograman <i>Python 3.7</i> ya sekarang.... Untuk masalah <i>machine learning library</i> ... ada <i>Tensorflow</i> ... dan <i>Pytorch</i>.... untuk <i>interface</i> sendiri kita ada <i>Falcon API framework</i> dari <i>Python</i> untuk <i>HTTP</i>.... Kalau di <i>tiket.com tools</i> untuk <i>meng-automate model cycle</i> adalah <i>Kubeflow</i>. (YP-4) Di kita pakai <i>Jupyter Notebook</i>.... (YP-5)</p>	<p>“Kal au bahasa aku pakai <i>Python</i>, <i>library</i> untuk mengolah datanya aku pakai <i>Pandas sama Numpy</i>. Kalau untuk visualisasi data pakai <i>Tableau</i>.” (TA-3) “Mu ngkin karena terbiasa aja sih pak, nyaman pakai <i>Visual Studio Code</i> nya. Sama pakai <i>Jupyter Notebook</i>, kalau untuk mengolah data yang bener-bener itu pakai <i>Jupyter Notebook</i>.” (TA-5)</p>	<p>“.... Karena kebanyakan permasalahan optimization kami pakai <i>Python, Numpy, Scipy, OR tools dari Google</i>. Tapi kalau ada pertanyaan apakah ada opportunity untuk pakai tools lain, ya kita sekali lagi tergantung kebutuhan. Kalau misalnya kami butuh <i>deep learning</i>, ya kami ambil <i>Pytorch</i> mungkin. (DR-12)</p>	<p>“...Nah, untuk <i>Deep Dive Analytics</i>-nya kita memakai si <i>Tableau</i>. ... seandainya mereka mau yang tidak berbayar, kita punya yang namanya <i>Data Studio</i>, itu juga <i>on by google</i> sama si <i>Metabase</i>.... Jadi kita akan pakai si <i>Tableau, Data Studio sama si Metabase</i>.” (AS-10) “... untuk <i>e-commerce</i> ... itu untuk permodelann ya dibangun pakai <i>Python</i>” (AS-14)</p>	<p>“... untuk tahap <i>EDA</i>, kita masih pakai di <i>Jupyter Notebook</i>. ... Tapi sebenarnya di kita itu ada <i>notebook</i> yang tersedia di <i>Kubernetes</i>.” (LW-4) “Sama sih, pak. Jadi memang kalau di saya untuk <i>EDA</i> dan <i>modelling</i> itu sama. Saya sama-sama menggunakan <i>python</i>.” (LW-5)</p>	<p>“Kalau untuk <i>tools</i> sebetulnya, kebetulan kalau saya lebih seringnya memang pakai <i>Python</i> ya pak....” (F-3) “Untuk yang visualisasi itu salah satunya kita bisa pakai <i>Tableau</i> ya, pak.... Walaupun misalkan kalau di <i>Python</i> kan memang ada, kita pakai <i>PyBlock</i> ataupun yang lain.” (F-12)</p>	<p>Semua Narasumber menggunakan bahasa pemrograman yang sama, yaitu <i>Python</i>. Tools yang biasa dipakai adalah <i>Falcon, Kubeflow, Jupyter Notebook, Tableau, Visual Studio Code, Numpy, Scipy, OR, Pytorch, Data Studio, Metabase, Kubernetes, dan PyBlock</i>. Jika diambil kesimpulan, maka tools yang paling banyak dipakai adalah <i>Jupyter Notebook, Numpy, dan Tableau</i>.</p>
--	--	---	---	--	---	---	--

Tabel 3 Tabel Triangulasi *Technology-Task Fitness*

AA	YP	TA	DR	AS	LW	F	Kesimpulan
<p>“... mungkin lebih ke kayak kecepatannya sih. Nah kita kan ada satu tahapan yang kayak proses transaksi gitu, yang tidak bisa ditulis dalam satu framework.” (AA-11)</p>	<p>“Kalau alat ukur model, kita kan model ada 2 yang diukur. Jadi yang model quality masalah akurasi, <i>precision, recall</i>.... Lalu yang kedua untuk mengukur kekuatan atau tahan banting.</p>	<p>“Biasanya kalau model itu kita lihat dari <i>confusion matrix</i>nya. <i>Positive true nya, positive false</i> sama <i>negative true, sama negative false</i>. Itu untuk model yang modelnya klasifikasi. Kalau untuk</p>	<p>“Pasti ada <i>feedback dari client</i> kan. Mungkin hampir semua project data science, kita kayak nggak bisa sekali, dua kali gitu kan. Jadi kayak prosesnya iterative kayak yang tepung itu</p>	<p>“... jadi kita pemilihan teknologynya dari sisi <i>data warehousing</i> adalah waktu itu yang jelas, satu harus <i>govern, governance-nya</i> harus jelas. Kemudian kita memilih suatu produk</p>	<p>“Nah, jadi <i>tim machine learning engineer</i> memastikan bahwa si model ini bisa di integrasi dengan baik, dengan sistem yang lain, seperti itu. Nah, biasanya mereka yang harus juga diskusi</p>	<p>“... kenapa memakai itu, karena memang dia itu satu, <i>open source</i>, kemudian dia juga lebih fleksibel ketika dia akan di <i>combine</i> dengan yang lain dan juga, ketersediaan salah</p>	<p>Kesulitan yang ditemukan oleh Peneliti, yaitu tentang kecepatan beberapa proses yang tidak bisa ditulis dalam satu framework. Di sisi lain pengukuran tingkat keberhasilan</p>

<p>“Nah ini, kalau untuk <i>anomaly detection</i> itu agak susah ... karena kan dia kan <i>unsupervised learning</i> kan.” (AA-12)</p>	<p>Itu kan kayak simulasi user, kita bisa set 1000 <i>user hit</i> dengan durasi 2 menit, apakah tahan. Kalau lolos yaudah.” (YP-9)</p> <p>“.... kalau user terima yasudah.... Kalau SLA nya dibawah 10 ms dihajar ternyata tidak 10 ms, nggak bisa kita harus kita harus develop lagi karena hubungannya sama <i>user</i>.” (YP-10)</p>	<p>regresi kita lihat error rate nya seberapa jauh.” (YP-18)</p>	<p>sudah versi ke-4. Jadi keberhasilan nya apakah kita berhasil nge-reduce cost gitu, dari berapa persen setahun. Jadi sekali lagi keberhasilannya sesuai dengan apa ekspektasi permasalahan annya dan masuk akal apa nggak. Kalau misalnya dalam setahun, kayanya bisa nih nge-reduce costnya 5% aja berarti kita berhasil.” (DR-19)</p>	<p>yang dimana bisa <i>scalable easily</i> dan juga bisa punya, itu tadi, balik lagi ke <i>governance</i>, pemotongan-pemotongan yang jelas. Jadi rumah-rumahnya masing-masing, jadi kita saling menjaga supaya masing-masing itu tidak saling kontek-mencontek antara data mereka.” (AS-6)</p>	<p>dengan tim infra, terus menghitung berapa lama, sih, satu proses itu akan selesai ... karena kan ada SLA yang harus kita jaga, gitu ya. Seperti kemarin itu kita pernah punya pengalaman, proses <i>image</i> terlalu berat, sehingga tidak lolos, nih, di quality, ketika di evaluasi oleh tim, gitu.” (LW-12)</p>	<p>satunya community yang menggunakan tools tersebut. Sehingga, kalau ada permasalahan apapun kita bisa <i>sharing</i> gitu dengan yang lain, sehingga lebih cepat.” (F-3)</p>	<p>n bisa diukur dari akurasi, kekuatan, <i>error rate</i>, <i>confusion matrix</i>, dan yang utama adalah <i>feedback</i> dari <i>client</i>. Hal tersebut bisa didapatkan melalui simulasi <i>user</i>, diskusi tim, dan evaluasi. Pemilihan <i>tools</i> berdasarkan kesesuaian dengan <i>governance</i>, tingkat kemudahan, fleksibilitas, ketersediaan komunitas, dan juga kelengkapan <i>library</i>.</p>
--	---	---	--	---	---	---	---

Tabel 4 Tabel Triangulasi Technology Acceptance Model

AA	YP	TA	DR	AS	LW	F	Kesimpulan
<p>“Ya karena kebiasaan kuliah sih sebenarnya, kebiasaan waktu kuliah aja. Kalau kita buka <i>interface</i> nya R itu kan sudah kelihatan ya R is for Statistical computing. Beda sama apa itu kalau Python kan untuk <i>general purpose</i> jadi bisa untuk bikin web dan yang lain.” (AA-9)</p>	<p>“Kalau R sendiri saya belum pernah coba sih, tapi kalau menurut yang saya baca dan saya dengar – dengar itu lebih cocok buat statistik. <i>Statistic base programing</i>, jadi kalau misal mau cari program regression itu biasanya pakai R.” (YP-18)</p>	<p>“Kalau mau gampang lebih baik pakai Azure pak atau orange3 itu untuk tools membuat model dan Analisa datanya itu tinggal drag and drop. Kalau mau yang lebih advanced sampai <i>reduce cost</i> juga, banyak ngodingnya itu lebih baik pakai Python.” (TA-16)</p>	-	<p>“... mungkin karena background-nya saya itu kalau di data itu sudah terlalu strong dimana anak-anak itu bisa melakukan <i>SQL programing</i> sendiri dan bisa dibilang sebagai <i>analytics</i> itu mereka, they're doing programing SQL-nya itu sudah advanced, cukup <i>advanced</i> lah untuk</p>	<p>“Tapi kalau dari saya pribadi, memang lebih senang menggunakan <i>Python</i> karena mungkin lebih familiar juga, terus untuk pengembangan ataupun untuk eksploratorin ya itu lebih fleksibel, gitu. Kita mau, misalnya cari distribusi datanya, ataupun mau lihat <i>outliers</i>, ataupun</p>	<p>“Iya dan orang lain juga bisa lebih mudah memahaminya, karena tujuannya di visualisasi itu kan bagaimana kita Insight-nya itu mereka bisa langsung <i>ngesh</i>, karena memang visualisasi itu setiap orang memahaminya gak mungkin sama. Itu yang kenapa orang sudah melihat awalnya saja</p>	<p><i>Tools</i> yang digunakan menjadi lebih mudah dipelajari karena sudah terbiasa sehingga cukup menguasai dan kesesuaiannya dengan kebutuhan. Beberapa Narasumber menyarankan untuk menggunakan R jika berhubungan dengan <i>Statistical Computing</i>, menggunakan <i>Azure</i></p>

				keperluan bisnis. (AS-12)	korelasi data itu, buat saya, sih, lebih mudah di <i>Python</i> , gitu, karena <i>library Python</i> juga cukup lengkap ya, untuk hal-hal yang berhubungan dengan eksploratori data analisis itu sendiri, sih.” (LW-3)	sudah gak menarik banget, itu sudah pasti mau <i>story telling</i> secantik apapun, mereka juga tidak akan begitu tertarik. Nah, itu kan juga salah satu faktor yang harus kita lihat juga.” (F-13)	atau <i>Orange3</i> untuk membuat model dan Analisa, dan menggunakan <i>Python</i> untuk programmming yang <i>advanced</i> . Dalam hal ini visualisasi menjadi hal yang penting karena orang lain akan mudah memahami data yang telah dibuat.
--	--	--	--	---------------------------	--	---	---

Dari ketiga tabel di atas tabel 2-4 dapat diambil kesimpulan dari masing-masing tabel triangulasi. Dari data Tabel 2-Demografi-Task dapat diambil kesimpulan bahwa narasumber memiliki pekerjaan dan tugas yang berbeda dan masih dalam kategori pekerjaan Data Science. Tugas mulai dari research untuk mendapatkan algoritma yang tepat, *mendevlop* model supaya bisa dipakai user, membuat machine learning, mengolah data sebagai dasar pengambilan keputusan, membuat teknologi lagi supaya data bisa dianalisis, melakukan eksperimen sebelum masuk ke machine learning, melakukan mapping data, melakukan standardisasi, membuat metadata, cleaning data, dan membuat visualisasi data.

Dari Tabel 2- Demografi-Tools dapat diambil kesimpulan bahwa semua Narasumber menggunakan bahasa pemrograman yang sama, yaitu Python. Tools yang biasa dipakai adalah Falcon, Kubeflow, Jupyter Notebook, Tableau, Visual Studio Code, Numpy, Scipy, OR, Pytorch, Data Studio, Metabase, Kubernetes, dan PyBlock. Untuk mayoritas tools yang paling banyak dipakai adalah Jupyter Notebook, Numpy, dan Tableau.

Dari sisi *Technology Task Fitness*, maka berdasarkan pada tabel 3, kesulitan yang ditemukan oleh Peneliti, yaitu tentang kecepatan beberapa proses yang tidak bisa ditulis dalam satu framework. Di sisi lain pengukuran tingkat keberhasilan bisa diukur dari akurasi, kekuatan, error rate, confusion matrix, dan yang utama adalah feedback dari client. Hal tersebut bisa didapatkan melalui simulasi user, diskusi tim, bantuan

komunitas dan evaluasi. Pemilihan tools berdasarkan kesesuaian dengan *governance*, tingkat kemudahan, fleksibilitas, ketersediaan komunitas, dan juga kelengkapan library.

Terakhir pada sisi *Technology Acceptance Model*, maka pada tabel 4 Tools yang digunakan menjadi lebih mudah dipelajari narasumber karena bila sudah terbiasa sehingga cukup menguasai dan menyesuakannya dengan kebutuhan. Beberapa Narasumber menyarankan untuk menggunakan R jika berhubungan dengan Statistical Computing, menggunakan Azure atau Orange3 untuk membuat model dan Analisa, dan menggunakan Python untuk programmming yang *advanced*. Dalam hal presentasi hasil, visualisasi menjadi hal yang penting karena orang lain akan mudah memahami data yang telah dibuat.

4 Kesimpulan

Teknologi pendukung *data science* masing-masing industri unggulan Surabaya dan sekitarnya berbeda-beda. Hal tersebut didasarkan pada beberapa aspek, seperti Demografi tugas dan tanggung jawab yang diberikan, Aspek Technology-Task Fitness, dan Aspek Technology Acceptance Model. Alasan utama penggunaan suatu *tools* adalah kemampuan pengguna dalam menguasai *tools* tersebut dan juga kesesuaiannya dengan *task*. Melalui penelitian ini, Penulis dapat menyimpulkan bahwa *tools* yang menjadi andalan industri di Surabaya dan sekitarnya adalah Jupyter Notebook, Python, dan Tableau. Keberhasilan dalam penggunaan *tools* tersebut tetap bergantung

pada kemampuan pengguna, hasil simulasi *user*, diskusi tim, dan proses evaluasi.

Referensi

- Ahmed, T., Chandran, V.G.R., Klobas, J.E., Liñán, F. and Kokkalis, P. (2020) 'Entrepreneurship education programmes: how learning, inspiration and resources affect intentions for new venture creation in a developing economy', *The International Journal of Management Education*, Vol. 18, No. 1, p.100327.
- Berman, F., Rutenbar, R., Hailpern, B., Christensen, H., Davidson, S., Estrin, D., Franklin, M., Martonosi, M., Raghavan, P., Stodden, V., & Szalay, A. S. (2018, April). Realizing the Potential of Data Science. *Communications of the ACM*, 61(04), 67–72. <https://doi.org/10.1145/3188721>
- Bischoff, K., Volkman, C. K., & Audretsch, D. B. (2018). Stakeholder collaboration in entrepreneurship education: An analysis of the entrepreneurial ecosystems of European higher educational institutions. *The Journal of Technology Transfer*, 43(1), 20–46.
- Liñán, F., Rodríguez-Cohard, J. C., & Rueda-Cantucho, J. M. (2011). Factors affecting entrepreneurial intention levels: A role for education. *The International Entrepreneurship and Management Journal*, 7(2), 195–218.
- Medeiros, M. M. de, Hoppen, N., & Maçada, A. C. G. (2020). Data science for business: benefits, challenges and opportunities. *Bottom Line*, 33(2), 149–163. <https://doi.org/10.1108/BL-12-2019-0132>
- Nabi, G., Walmsley, A., Liñán, F., Akhtar, I., & Neame, C. (2018). Does entrepreneurship education in the first year of higher education develop entrepreneurial intentions? The role of learning and inspiration. *Studies in Higher Education*, 43(3), 452–467.
- O. Nyumba, T., Wilson, K., Derrick, C. J., & Mukherjee, N. (2018). The use of focus group discussion methodology: Insights from two decades of application in conservation. *Methods in Ecology and Evolution*, 9(1), 20–32.
- Rogala, P., Batko, R., & Wawak, S. (2017). Factors affecting success of training companies. *Studies in Continuing Education*, 39(3), 357–370. <https://doi.org/10.1080/0158037X.2017.1336995>
- Tanamal, R. (2019). What Is the Most Influential Factor On Decisions Using Youtube As A Tool To Support Buy Or Sell Means? Case Study Surabaya City And Surrounding Area. *Journal of Theoretical and Applied Information Technology*, 97(20), 2406-2418.
- Wiradinata, T., & Antonio, T. (2018). Developing Technology Entrepreneurship Subjects: A Four-Year Evaluation. *International Journal of Innovation, Management and Technology*, 9(1), 37–41. <https://doi.org/10.18178/ijimt.2018.9.1.784>.
- Wu, Y. C. J., Kuo, T., & Shen, J. P. (2013). Exploring social entrepreneurship education from a web-based pedagogical perspective. *Computers in Human Behavior*, 29(2), 329–334.