

Kebenaran dalam Perspektif Filsafat Ilmu Pengetahuan dan Implementasi dalam Data Science dan Machine Learning

Mohammad Idris¹, Riza Ibnu Adam², Yulrio Brianorman³, Rinaldi Munir⁴, Dimitri Mahayana⁵
^{1,2,3,4,5} Sekolah Tinggi Elektro dan Informatika, Institut Teknologi Bandung, Indonesia
E-mail: 33221010@std.stei.itb.ac.id¹, 33221048@std.stei.itb.ac.id²,
33221012@std.stei.itb.ac.id³, rinaldi-m@stei.itb.ac.id⁴, dimitri@lskk.ee.itb.ac.id⁵



This is an open-access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.
Copyright © XXXX by Author. Published by Universitas Pendidikan Ganesha.

Diterima: 15-12-2021

Direview: 15-12-2021

Publikasi: 30-06-2022

Abstrak

Perkembangan *data science* atau *machine learning* mengalami perkembangan yang sangat cepat. Perkembangan ini berpotensi menjadi suatu *pseudoscience* akibat proses pengumpulan, pemakaian algoritma serta proses pengolahan data yang tidak sesuai dengan standar baku penelitian. *Machine learning* adalah satu dari sekian teknik di bidang artificial intelligence (AI) yang memungkinkan suatu mesin untuk belajar mandiri. Beberapa implementasi *machine learning* seperti robot konsultan, pengoptimalan proses, penilaian kredit, bidang keamanan, serta pelayanan. Algoritma *supervised learning* merupakan salah satu yang sering dipakai pada bidang data science. Penerapan algoritma *supervised learning* dalam dunia *data science* bisa terjebak menjadi *pseudoscience* akibat adanya kesalahan dalam pemakaian dan pengolahan data pada penelitian. Solusi untuk menghindarinya adalah dengan menerapkan konsep epistemologi Karl Popper yaitu berkaitan dengan falsifikasi sains untuk menyelesaikan persoalan demarkasi. Untuk memperkuatnya dapat juga menggunakan prinsip 4 teori kebenaran yaitu koherensi, korespondensi, pragmatis dan konsensus.

Kata Kunci: *data science*; demarkasi; falsifikasi; *machine learning*; *pseudoscience*

Abstract

The development of data science or machine learning is experiencing swift growth. This development can become a pseudoscience due to the process of collecting, using algorithms, and processing data that are not per research standards. Machine learning is one of numerous artificial intelligence (AI) techniques that allow a machine to learn independently. Several machine learning implementations include consulting robots, process optimization, credit scoring, security, and services. Data science often uses supervised learning algorithms, which can be trapped into pseudoscience due to errors in the use and processing of data in research. The solution to avoid this is to apply the epistemological concept of Karl Popper, which is related to the falsification of science to solve the problem of demarcation. To enhance it, you can also use the principles of the Four Theory of Truth, namely coherence, correspondence, pragmatism, and consensus.

Keywords: data science; demarcation; falsification; machine learning; pseudoscience

1. Pendahuluan

Ilmu data atau *machine learning* berkembang sangat pesat. Namun, secara umum, ini dapat dibagi menjadi tiga zaman (Suyanto, 2018). Artinya, zaman pra-1980 di mana hampir semua metode pembelajaran mesin diselidiki dalam zaman ini untuk menghasilkan permukaan penentuan linier. Metode pembelajaran linier ini sudah memiliki dasar pemikiran yang kuat. Pada akhir 1980-an, pohon keputusan dan Jaringan Syaraf Tiruan mempelopori pembelajaran nonlinier. Namun, alasan itu masih lemah. Selain itu, kedua metode tersebut sering kali menggunakan optimasi lokal. Era dari tahun 1990-an hingga sekarang. Selama periode ini, metode pembelajaran nonlinier yang efisien dikembangkan berdasarkan teori pembelajaran komputer. Metode pembelajaran nonlinier ini memiliki dasar pemikiran yang kuat.

Ilmu data adalah kombinasi dari beberapa bidang yang membantu untuk memproses data dan memperoleh informasi yang berguna, seperti matematika, statistik, dan komputer (Kumar, Dhanda, & Pandey, 2018). Ilmu data menganalisis data dari kumpulan data kecil (sampel) dan besar (populasi) (analisis data) dengan menerapkan algoritma khusus yang bertujuan mengekstraksi data (penambangan data) dan memperoleh pola data. Menggabungkan matematika, statistik, dan ilmu komputer dengan tujuan menciptakan kecerdasan sistem (AI) yang dapat memprediksi data dengan akurasi yang cukup (prediksi), membantu untuk membuat keputusan, dan membuat untuk terus belajar, sehingga dapat digunakan untuk (pembelajaran mesin).

Tujuan dari ilmu data adalah untuk mempelajari data dan memberikan informasi seakurat mungkin. Ini digunakan untuk tujuan berikut (Jimy, 2019).

- a. Deskripsi yang menunjukkan sampel data untuk analisis dan penemuan masalah.
- b. Membuat prediksi, yaitu membuat prediksi dalam bentuk nilai, probabilitas, dan data, serta merekomendasikan hasilnya sebagai pendukung keputusan atau langsung oleh sistem. Sains secara harfiah berarti "mengetahui".

Di bawah ini, adalah tiga pernyataan tentang sains. Pertama, sains adalah pengetahuan yang diperoleh melalui pengalaman. Kedua, sains adalah kumpulan pengetahuan tentang alam dan materi melalui pengamatan dan eksperimen. Bagaimanapun, sains adalah fakta atau prinsip yang diturunkan dari studi sistem.

Untuk melakukan sains, kita harus berasumsi bahwa ada keteraturan di alam semesta. Oleh karena itu, pikiran manusia dapat memahami tatanan ini. Jika kondisinya sama, hasil setiap tes akan sama. Tidak peduli berapa banyak percobaan yang dilakukan, maka akan selalu mendapatkan hasil yang sama. Namun, jika eksperimen menghasilkan hasil yang berbeda, hal itu disebabkan oleh kesalahan antara eksperimen saat ini dan sebelumnya.

Sains palsu atau *pseudoscience* adalah pengetahuan yang terdiri dari pernyataan, kepercayaan, atau praktik yang mengklaim sebagai sesuatu yang ilmiah dan faktual tetapi tidak sesuai dengan metode ilmiah. *Pseudoscience* sering dicirikan oleh klaim berlebihan yang kontradiktif, ketergantungan pada bias konfirmasi daripada argumen tandingan, kurangnya keterbukaan terhadap evaluasi oleh para ahli lain, dan kurangnya praktik sistematis dalam membuat hipotesis. Bergantung pada data yang digunakan dalam penelitian, ilmu data dan pembelajaran mesin dapat menjadi pseudosains. Faktor-faktor yang berkontribusi terhadap hal ini termasuk kurangnya pemahaman tentang masalah di area di mana solusi dicari, metode pengumpulan data yang tidak standar, persiapan data untuk diproses melalui tahapan yang tidak standar, dan model atau algoritma yang digunakan. Kesalahan dalam pemilihan, evaluasi / metode pengujian yang tidak dilakukan dalam proses, evaluasi/pengujian sesuai kebutuhan, dan kesulitan pengguna model untuk digunakan. Ada beberapa kesalahan yang terjadi dalam proyek *data science* (Olavsrud, 2020). Public Health England (PHE) telah kehilangan ribuan kasus COVID karena telah melampaui batas data spreadsheet. PHE menggunakan pendekatan otomatis untuk mentransfer hasil tes positif COVID19 dalam format file CSV ke template Excel yang digunakan untuk dasbor laporan dan pelacakan kontak. Tabel Excel dapat berisi hingga 1.048.576 baris dan 16.384 kolom per lembar kerja. PHE kemudian mendaftarkan kasus dalam kolom, bukan baris. Jika kasus melebihi maksimum 16.384 kolom, Excel memotong 15.841 entri terbawah.

Menurut sebuah penelitian yang diterbitkan di Science pada 2019, algoritma prediksi perawatan kesehatan yang digunakan oleh rumah sakit dan perusahaan asuransi AS untuk mengidentifikasi pasien yang membutuhkan program 'manajemen perawatan risiko tinggi' dapat memilih orang kulit hitam. Ternyata jenis kelaminnya jauh lebih rendah. Program Manajemen Perawatan Risiko Tinggi memberi pasien penyakit kronis staf perawat terlatih dan pengawasan dokter keluarga untuk menghindari konsekuensi serius. Namun, algoritma itu jauh lebih mungkin untuk merekomendasikan pasien kulit putih ke program ini daripada pasien kulit hitam. Alat perekrutan berbasis AI Amazon hanya menyarankan pria. Amazon, seperti banyak perusahaan besar lainnya, sedang mencari alat untuk membantu departemen pengembangan sumber daya menyaring pelamar terbaik. Untuk itu, Amazon mulai mengembangkan perangkat lunak rekrutmen bertenaga AI pada tahun 2014. Hanya ada satu masalah. Sistem ini memprioritaskan calon laki-laki. Reuters melaporkan bahwa Amazon telah menghentikan inisiatif tersebut pada 2018.

Berdasarkan uraian diatas maka, suatu penelitian di bidang Data Science dan Machine Learning jika tidak memperhatikan dasar-dasar pengetahuan yang tepat dan dilakukan dengan baik dapat berubah menjadi pseudoscience. Untuk mencegah hal ini diterapkan suatu sistem

demarkasi, yaitu memberlakukan sistem saintifik yang dapat dilakukan proses falsifikasi berdasarkan pengalaman sesuai premis dari Popper. Melalui demarkasi popper, dapat diketahui bahwa suatu sistem dapat dikatakan sebagai sesuatu yang bersifat saintifik jika hal tersebut bisa dilakukan pengujian. Dalam hal ini, pengujian yang dimaksud adalah dapat diuji dengan cara apakah sesuatu tersebut mempunyai suatu kemungkinan difalsifikasi berdasarkan pengalaman. Untuk memperkuatnya dapat juga menggunakan prinsip 4 teori kebenaran yaitu koherensi, korespondensi, pragmatis dan konsensus.

2. Metode

Bagian ini membahas epistemologi Karl Popper, masalah induksi, masalah demarkasi dan tampering. Ketiga hal tersebut membentuk epistemologi Popper dan saling terkait. Ini juga menggambarkan empat teori kebenaran: koherensi, korespondensi, kepraktisan, dan konsensus.

3. Hasil dan Pembahasan

a. Epistemologi Karl Popper

1) Persoalan Induksi

Induksi adalah cara untuk menentukan kebenaran universal melalui pengamatan terhadap objek tertentu. Menarik kesimpulan umum saja dengan menggunakan unsur-unsur khusus atau unsur-unsur tertentu. Metode ini sangat mirip dengan ilmu empiris (Rahman, 2017). Oleh karena itu, menurut Popper, pandangan seperti itu tidak dapat dibenarkan. Dia berpendapat bahwa proposisi bisa menjadi proposisi universal. Ini sangat tidak logis. Menurutnya, jumlah angsa putih yang ditemukan tidak dapat membenarkan kesimpulan bahwa ini semua adalah angsa putih. Dengan kata lain, pemikiran induktif ini tidak dapat dibenarkan secara logis. Popper, di sisi lain, mengakui bahwa pengalaman empiris adalah cara untuk menentukan kebenaran. Jelas bahwa metode induktif membantu menentukan probabilitas daripada kebenaran. Namun, prinsip induksi dengan pengamatan berulang dapat menyebabkan kemunduran. Ada siang dan malam setiap hari dari masa lalu hingga hari ini, tetapi tidak dapat ditentukan bahwa hal yang sama akan terjadi. Ini karena tidak ada kepastian logis bahwa ini akan dikatakan.

2) Persoalan Demarkasi

Demarkasi adalah batas-batas yang ditetapkan oleh para pejuang (pihak perang) untuk memisahkan dua pasukan yang berlawanan [6]. Positivisme berpendapat bahwa induksi dapat digunakan sebagai prinsip verifikasi. Kajian ini bisa menjadi dasar ilmu atau referensi ilmiah. Hal ini memerlukan observasi, induksi, dan verifikasi agar ilmu dapat diklasifikasikan sebagai penting (penting) atau tidak penting (tidak penting) (Dochmie, 2018). Pendapat ini ditentang oleh Popper bahwa topik ini tidak bermakna atau tidak berarti, tetapi merupakan topik ilmiah dan non-ilmiah. Seperti dijelaskan sebelumnya, Popper menolak menggunakan validasi yang menekankan pengamatan inderawi induktif sebagai standarisasi ilmu dan pengetahuan ilmiah (pemisahan). Pada dasarnya batas adalah batas-batas yang dibentuk oleh positivisme yang berkaitan dengan pengetahuan ilmiah (sains) dan pseudosains (*pseudoscience*). Popper mengusulkan kriteria batas. Ini berarti bahwa suatu sistem hanya dapat digambarkan sebagai ilmiah jika dapat diperiksa, yaitu jika dapat dirusak oleh pengalaman. Menurut Popper, falsifiability of experience merupakan kriteria untuk membedakan antara sains dan non-sains (Mahayana, 2018).

3) Persoalan Demarkasi

Karl Popper memecahkan masalah induksi dengan merumuskan kembali pertanyaan tersebut secara lengkap. Popper percaya bahwa semua metode ilmiah tidak ada hubungannya dengan induksi (Nielson, 2021). Bagi Popper, parameter sains adalah argumen tandingan, bukan verifikasi. Artinya jika teori tersebut dapat dikritisi, maka dapat disebut ilmiah (Nur, 2012). Teori yang secara khusus menyatakan mengesampingkan kemungkinan kepalsuan bukanlah ilmu pengetahuan. Pengujian lebih lanjut dan menjelaskan kesalahan-kesalahan yang dapat terjadi dalam teori memiliki pengertian bahwa logika teori menjadi lebih valid dan teruji. Dalam kasus popper, pengetahuan teori tidak harus dibuktikan (bukti kebenaran), tetapi harus diuji untuk kemungkinan kesalahan. Pengetahuan telah bertahan dari pemeriksaan dan dapat dikatakan terbukti menjadi karakteristik atau ditingkatkan berdasarkan pengalaman masa lalu, kecuali jika digantikan oleh penemuan-penemuan lain dalam kemajuan ilmiah.

b. Empat Teori Kebenaran

1) Teori Kebenaran Koherensi

The Coherentism Theory of Truth adalah teori kebenaran yang didasarkan pada standar kebenaran. Suatu pernyataan dikatakan benar jika konsisten atau konsisten dengan pernyataan yang telah terbukti sebelumnya.

Satu pernyataan tidak secara logis bertentangan dengan yang lain. Misalnya, semua manusia membutuhkan oksigen, dan Amir adalah manusia, maka Amir membutuhkan oksigen. Agar dianggap benar, teori ini mensyaratkan bahwa tidak ada konsistensi atau kontradiksi antara pernyataan dan aksioma. Untuk itu teori koherensi disebut juga teori koherensi (Padli & Mustofa, 2021).

2) Teori Kebenaran Korespondensi

Teori kebenaran korespondensi, kadang-kadang disebut , Correspondence Theory of Truth, adalah teori bahwa suatu pernyataan benar jika cocok dengan fakta atau pernyataan yang ada di alam, atau objek yang menjadi subjek pernyataan (Atabik, 2014; Idaman, 2019). Misalnya, sekolah memiliki 30 kendaraan yang diparkir. Pembuat pernyataan menghitung dirinya sendiri setelah melihatnya dengan panca inderanya.

Teori kebenaran korespondensi membutuhkan pengalaman sebagai dasar kebenaran. Kebenaran ilmiah perlu dikaitkan dengan teori kebenaran korespondensi. Kita tidak dapat melihat gravitasi, tetapi kita dapat mengamati peristiwa yang diprediksi oleh gravitasi.

3) Teori Kebenaran Pragmatis

Teori kebenaran praktis adalah teori kebenaran yang paling sederhana. Menurut teori ini, suatu pernyataan atau gagasan dianggap benar jika bermanfaat atau tersedia bagi banyak orang. Teori ini tidak ada hubungannya dengan korelasi atau korespondensi. Selama sesuatu itu berguna bagi manusia, itu berarti sesuatu itu benar.

Misalnya, kalimat "semua besi memuai jika dipanaskan" adalah alat yang berguna bagi manusia karena merupakan kebenaran praktis bagi pandai besi. Pandai besi tidak membutuhkan bukti empiris bahwa besi memuai.

4) Teori Kebenaran Konsensus

Teori kebenaran konsensus awalnya ditemukan oleh sejarawan sains Thomas Kuhn. Para penulis *The Structure of Scientific Revolutions* yang menyatakan bahwa sains berkembang dalam beberapa tahap (Padli & Mustofa, 2021). Pertama, sains bisa menjadi sains biasa jika diterima oleh masyarakat atas dasar konsep kebenaran ilmiah. Kedua, ada beberapa anomali dalam perkembangannya yang meragukan validitas konsep kebenaran. Selain itu, ada revolusi ilmiah yang akan membawa perubahan paradigma ke komunitas ilmiah. Singkatnya, evolusi ilmu pengetahuan ditandai dengan perubahan paradigma lama yang digantikan oleh paradigma baru. Perubahan ditentukan oleh paradigma penerimaan sosial (*social acceptance*) dan konsep kebenaran ilmiah.

Menurut konsep Kuhn, teori ilmiah dianggap benar selama ada dukungan atau konsensus tentang kebenaran teori dalam komunitas ilmiah. Ini adalah teori kebenaran konsensus.

c. Jenis-Jenis *Machine Learning*

Machine learning adalah salah satu teknik yang digunakan dalam kecerdasan buatan untuk memungkinkan mesin belajar dari data tanpa diprogram secara eksplisit. Salah satu algoritma pembelajaran mesin yang dapat digunakan adalah jaringan saraf. Ini adalah program dengan banyak status bercabang yang dapat meniru kecerdasan manusia seperti sistem pakar, tetapi tidak termasuk pembelajaran mesin karena program tidak belajar dari data dan secara eksplisit ditulis oleh programmer. Evolusi pembelajaran mesin menarik gaya dan jenis yang berbeda. Tentunya jenis *machine learning* ini memiliki ciri dan ciri tersendiri (Ray, 2019).

Algoritma pembelajaran terawasi adalah algoritma yang paling banyak digunakan di dunia untuk ilmu data dibandingkan dengan pembelajaran tanpa pengawasan. Analisis regresi multilinier dan logistik, adalah contoh dari pembelajaran terawasi. Perbedaan kedua algoritma tersebut adalah bagaimana mempelajari untuk memprediksi dan mengklasifikasikan.

Dalam pembelajaran yang diawasi, algoritma tampaknya telah dilatih sebelumnya untuk memungkinkan prediksi dan klasifikasi. Implementasi dari algoritma supervised learning dalam

dunia *data science* adalah support vector machine, neural network, regresi linier dan logistik, random forest, dan pohon klasifikasi. Metode unsupervised learning tidak menggunakan data latih dan data latih untuk melakukan prediksi dan klasifikasi. Berdasarkan model matematis, algoritma ini tidak memiliki variabel target. Salah satu tujuan dari algoritma ini adalah untuk mengelompokkan objek yang hampir identik dalam area tertentu. Contoh pendekatan ini adalah ketika seorang analis data mengelompokkan pelanggan berdasarkan kesamaan pendapatan, usia, liburan, dan jenis pekerjaan. Tidak ada data pelatihan yang diperlukan untuk menentukan peringkat pelanggan berdasarkan karakteristik umum ini. Dengan demikian maka dapat mengumpulkan pelanggan ini secara langsung dengan menggunakan data yang ada. Algoritma yang dapat digunakan dalam pembelajaran tanpa pengawasan termasuk KMeans dan pengelompokan hierarkis.

Reinforcement learning (RL) berbeda dengan pembelajaran terawasi dan tak terawasi. Algoritma ini bertujuan agar komputer dapat belajar secara mandiri dari lingkungan (environment) melalui suatu agen. Komputer belajar secara mandiri dengan berinteraksi dengan lingkungan (self-discovery). Pembelajaran penguatan didefinisikan sebagai metode pembelajaran mesin yang berhubungan dengan bagaimana agen perangkat lunak berperilaku di lingkungan. Ini adalah bagian dari pendekatan pembelajaran mendalam yang membantu memaksimalkan penghargaan kumulatif. Fitur belajar gain tidak termasuk pemantauan, hanya bilangan real atau sinyal. Pengambilan keputusan berurutan. Waktu memainkan peran penting dalam membuat segalanya lebih kuat. Umpam balik selalu tertunda. Tindakan agen menentukan data apa yang akan diterima selanjutnya.

d. Analisis dan Penerapan Metode

1) *Machine Learning* dalam Penyelesaian Sebuah *Use Case*

Dalam beberapa tahun terakhir, industri keuangan semakin memperkenalkan solusi cerdas untuk beradaptasi dengan perubahan kondisi industri. Kecerdasan buatan (AI) dan pembelajaran mesin (ML) sekarang banyak digunakan di berbagai industri. Pada 2019, kecerdasan buatan di pasar teknologi keuangan mencapai \$ 6,67 miliar. Jumlah ini diperkirakan akan meningkat melebihi \$22,6 miliar selama lima tahun ke depan. Tingkat pertumbuhan tahunan rata-rata (CAGR) untuk periode tersebut mencapai 23,37.% dan tidak ada tanda-tanda perlambatan. Saat ini, teknologi kecerdasan buatan digunakan untuk menyelesaikan berbagai tantangan bisnis. Penggunaan kecerdasan buatan dan pembelajaran mesin yang paling umum adalah penghematan biaya 38%, pengiriman informasi 37%, pengalaman pelanggan 34% lebih baik, otomatisasi proses internal 30%, deteksi penipuan 27% , 26. Termasuk% kepuasan pelanggan. Komunitas menemukan bahwa industri keuangan akan terus menawarkan layanan sesuai permintaan hingga 2021. Singkatnya, pelanggan memiliki layanan dan produk yang lebih cepat dan personal daripada sebelumnya. Berikutnya adalah implementasi *machine learning* yang sedang tren saat ini.

(a) Robot Konsultan

Sejalan dengan topik pendanaan dan kebutuhan, rekomendasi produk yang dipersonalisasi dan manajemen portofolio adalah dua dari dua solusi paling populer untuk AI dan pembelajaran mesin pada tahun 2021. Keakuratan dan standar etika teknologi kecerdasan buatan masih kontroversial, tetapi nilai dari aplikasi tersebut tetap ada. tumbuh besar. Lembaga keuangan yang menggunakan solusi terbaru dapat menawarkan peluang investasi kepada klien berdasarkan pendapatan, kebiasaan investasi saat ini, dan manajemen risiko.

(b) Pengoptimalan Proses

Salah satu penggunaan kecerdasan buatan dan pembelajaran mesin yang paling umum dalam teknologi keuangan adalah pengoptimalan proses, yang akan diperluas lebih lanjut pada tahun 2021. Optimalisasi proses membantu lembaga keuangan mengurangi beban kerja karyawan mereka. Untuk membuat seluruh proses lebih efisien dan produktif. Pengoptimalan proses biasanya digunakan untuk mengotomatisasi fungsionalitas pusat panggilan. Dokumentasi chatbot terkait pelanggan yang dioptimalkan dan pelatihan karyawan yang ditingkatkan.

(c) Penilaian Kredit

Beberapa sistem penilaian kredit saat ini sudah ketinggalan zaman. Sistem membuat keputusan berdasarkan demografi seperti pekerjaan, usia, ras, dan jenis kelamin, tetapi jarang memperhitungkan keterampilan dan risiko pemberi pinjaman. Kecerdasan buatan dan

pembelajaran mesin memungkinkan lembaga keuangan menjelaskan risiko pelanggan mereka dengan lebih akurat.

(d) Keamanan

Penipuan pada tahun 2020. Tiga perlima dari organisasi ini melaporkan peningkatan penipuan setiap tahun. Beberapa masalah yang paling mengkhawatirkan terkait dengan pembukaan rekening dan penipuan keuangan. Dan penipuan di sektor keuangan bukanlah hal baru. Digitalisasi berarti scammers dapat bertindak lebih kreatif. Oleh karena itu, penyedia jasa keuangan perlu meningkatkan keamanan.

(e) Layanan Pelanggan

Tidak ada keraguan bahwa pendapat pelanggan sangat penting. Tidak ada keraguan bahwa umpan balik menjadi sangat penting. 93% konsumen memilih lembaga keuangan yang menawarkan layanan pelanggan yang sangat baik. Jadi apa sebenarnya layanan pelanggan yang baik itu? Jawabannya dapat diringkas dalam dua poin: waktu respons dan layanan individual. Kecerdasan buatan dan pembelajaran mesin memainkan peran penting di sini. Menentukan bagaimana dan di mana menerapkan solusi baru bukanlah tugas yang mudah bagi perusahaan mana pun. Meskipun 84% pemimpin bisnis percaya bahwa kecerdasan buatan perlu digunakan untuk mencapai tujuan pertumbuhan perusahaan. Lebih dari 76% eksekutif tidak tahu bagaimana mencapai tujuan ini. Mengembangkan bisnis dilakukan dengan kecerdasan buatan dan pembelajaran mesin tidaklah mudah. Namun, ini adalah alat penting bagi organisasi yang bergerak maju. Jadi mari kita menjadikannya pengalaman pribadi. Membuat keputusan internal yang lebih cerdas. Memprediksi tren bisnis, interaksi pelanggan, atau menemukan pola target penipuan. Perusahaan perlu memperhatikan kebutuhan bisnis mereka dan menemukan kecerdasan buatan dan pembelajaran mesin yang tepat untuk bisnis mereka (Warin, 2021).

2) **Confirmation Bias di Seluruh Tahapan Perencanaan Machine Learning**

Confirmation bias dapat menyebabkan *data science* menjadi *pseudoscience*. Hal ini dikarenakan hal sebagai berikut.

- (a) Model *machine learning* dapat dianggap sebagai kelas program yang memiliki metode map untuk memetakan input menjadi output.
- (b) Model *machine learning* juga memiliki parameter dan hyperparameter yang nilainya memengaruhi map (dan output).
- (c) Nilai parameter dioptimisasi berdasarkan data latih agar diperoleh pemetaan input-output yang sesuai dengan data.
- (d) Nilai hyperparameter tidak bergantung pada data dan konstan selama model dioptimisasi. Nilai yang paling tepat untuk hyperparameter dipilih melalui hyperparameter tuning.

3) Penerapkan **Critical Thinking** Karl Popper

Tujuan ilmiah dari pendekatan pembelajaran mesin adalah untuk mengembangkan metode analisis data prediktif yang andal. Keandalan ini ditunjukkan dengan kemampuan model untuk memprediksi hubungan antara variabel eksternal dari data yang digunakan untuk memperkirakan model (dikenal sebagai validasi/pengujian data). Data yang digunakan untuk mengestimasi model disebut data latih. Oleh karena itu, para ilmuwan data saat ini sedang mencari cara terbaik untuk melakukan ini. Keunggulan dalam peramalan yang konsisten dan berkelanjutan. Ini membuka jalan bagi abstraksi pemikiran filosofis tentang apa yang sebenarnya terjadi di antara banyak variabel. Tujuannya adalah untuk menjawab pertanyaan tentang mengapa fenomena prediktibilitas dari sekumpulan variabel lebih unggul dari variabel lainnya.

Metode pengembangan ilmiah di atas adalah gaya pembelajaran mesin. Ini sangat berbeda dari pendekatan tradisional, model statistik. Dalam model statistik, peneliti memulai dengan pertanyaan penelitian atau hipotesis yang disintesis dari beberapa teori dalam literatur. Dari situ, model statistik dibuat dengan harapan akan memiliki kemampuan terbaik dan paling tepat untuk menjawab pertanyaan penelitian.

Untuk mengevaluasinya, kriteria berbeda digunakan. Yang pertama adalah signifikansi individu (uji), makna model, koefisien determinasi lengkap seperti keseragaman atau non-stasioneritas (jika datanya deret waktu). Pendekatan ini terjadi di banyak bidang, bukan hanya keuangan. Tentu saja, hanya karena satu set variabel dengan benar memprediksi perilaku variabel lainnya tidak berarti ada sesuatu yang mendasar di antara variabel-variabel tersebut. Francis mengatakan bahwa melihat perilaku hubungan antara variabel (X dan Y), selalu ada lima

kemungkinan: (a) X mempengaruhi Y, (b) Y mempengaruhi X. Memberi, (c) X dan Y saling mempengaruhi, (d) Y dan X dipengaruhi oleh variabel lain Z, dan (e) X dan Y tidak relevan. Pola a sampai e semuanya memberikan prediksi untuk model. Tapi secara teori, biasanya ada batasannya. Misalnya, logis bahwa X mempengaruhi Y, tetapi tidak sebaliknya.

Uraian singkat ini dimaksudkan untuk memberikan gambaran tentang nilai-nilai penting yang harus diberikan oleh *machine learning*. Ini adalah bidang yang "dilahirkan kembali" dan sangat menjanjikan sebagai alternatif pengembangan ilmu pengetahuan. Komunitas ilmiah (khususnya praktisi) terkenal, dan terutama karya monumental yang dihasilkannya, mendapat penghargaan khusus (Team, 2020).

4) Mengadopsi CRISP-DM

Bagian ini menjelaskan CRISPDM, yang merupakan singkatan dari Cross Industry Standard Process for Data Mining. Menurut banyak survei dan jajak pendapat pengguna, ini tetap menjadi standar de facto untuk membangun proyek penggalian data dan penemuan pengetahuan. Namun, selama dua dekade terakhir, area ini telah berkembang secara signifikan dan telah menjadi formulasi yang direkomendasikan untuk ilmu data sebagai alternatif penambangan data (Martínez-Plumed, Contreras-Ochando, Ferri, & Hernández-Orallo, 2021).

Business Understanding, ini merupakan langkah awal dalam CRISPDM dan memiliki bagian yang sangat penting. Langkah ini membutuhkan pengetahuan tentang objek bisnis, cara membuat atau mengambil data, dan bagaimana membandingkan tujuan pemodelan dengan tujuan bisnis untuk membangun model yang optimal untuk secara jelas menentukan tujuan umum.

Data Understanding, umumnya didefinisikan untuk menyelidiki data dan mengidentifikasi masalah dengan data. Langkah ini memberikan landasan analitis untuk penelitian dengan membuat ringkasan dan mengidentifikasi potensi masalah dengan data yang ada. Sekali lagi, langkah ini harus dilakukan dengan hati-hati, bukan terburu-buru, seperti visualisasi data. Visualisasi data dapat membingungkan bila dikombinasikan dengan data ringkasan. Jika masalah tetap tidak terselesaikan pada saat ini, prosedur pemodelan akan terganggu. Ringkasan atau ringkasan data dapat membantu untuk memastikan bahwa data yang disajikan sesuai dengan yang diharapkan, atau mengungkapkan bias yang tidak diharapkan yang perlu diatasi pada tahap selanjutnya, yaitu dalam penyusunan data. Masalah dengan data, seperti nilai yang hilang, outlier, distribusi spike, dan distribusi dua arah, perlu diidentifikasi dan diukur sehingga dapat diperbaiki selama persiapan data.

Data Preparation, pada langkah ini Variabel turunan dibuat untuk menyiapkan data, pada dasarnya untuk memperbaiki masalah data. Jelas, langkah ini membutuhkan pemikiran dan upaya untuk memastikan bahwa data sesuai dengan algoritma yang digunakan. Ini tidak berarti bahwa data dapat digunakan sampai langkah terakhir dalam operasi pra-data pertama yang memecahkan masalah data. Fase ini sering dipertimbangkan ketika mencari masalah selama pengembangan model. Ini diulang sampai ditemukan kecocokan untuk data. Langkah pengambilan sampel dapat dilakukan di sini dan data biasanya dibagi menjadi dua data pelatihan dan data uji. Kegiatan yang dilakukan meliputi pemilihan kasus dan parameter (seleksi data) untuk dianalisis, konversi (transformasi) dari parameter spesifik, dan pembersihan data.

Modelling, secara sederhana fase yang digunakan untuk membuat model prediktif atau deskriptif. Pada titik ini, gunakan teknik statistik dan pembelajaran mesin untuk menentukan teknik penambangan data, alat penambangan data, dan algoritma penambangan data yang akan digunakan. Langkah selanjutnya adalah menerapkan teknik dan algoritma penambangan data ke data menggunakan alat baru. Jika perlu menyesuaikan data tersebut untuk teknik penambangan data tertentu, Sehingga dapat kembali ke persiapan data. Beberapa model yang paling umum adalah klasifikasi, penilaian, peringkat, pengelompokan, penemuan hubungan, dan karakterisasi.

Evaluation, adalah fase yang menginterpretasikan hasil penambangan data yang dihasilkan selama proses pemodelan fase sebelumnya. Evaluasi model yang diterapkan pada langkah sebelumnya dilakukan dengan tujuan agar model yang ditentukan sesuai dengan tujuan yang dicapai pada langkah pertama.

Deployment adalah langkah terpenting dalam proses CRISPDM. Perencanaan implementasi dimulai pada pemahaman bisnis, tidak hanya bagaimana nilai model dihasilkan, tetapi juga bagaimana nilai keputusan dimodifikasi dan bagaimana keputusan diintegrasikan ke dalam sistem operasional. Akhirnya, rencana penyebaran sistem mengakui bahwa tidak ada yang namanya model statis. Model dibangun dari data yang diwakili oleh data pada waktu

tertentu, sehingga properti data dapat berubah seiring waktu. Model juga perlu dilacak dan dapat diganti.

5) Analisa

Dapat dikatakan bahwa kita berada di dunia hiper informasi di era ini. Kurangnya data untuk menarik kesimpulan tidak lagi menjadi masalah. Kita dapat memilih dari ribuan atribut data terkait, ataupun jutaan. Sederhananya, adalah perlu tahu cara mengapung di "lautan data". Menurut Ahrens, struktur data yang kompleks memiliki tiga karakteristik. Yaitu, (a) berdimensi tinggi, (b) data tinggi, dan (c) data besar. Data berdimensi tinggi adalah struktur data yang memiliki jumlah atribut yang banyak dan dapat melebihi jumlah observasi. Di era ini, internet dan media social, ada banyak atribut yang bisa didapatkan dari seseorang. Sepuluh tahun yang lalu, pemasaran dan survei lainnya mungkin memerlukan informasi seperti usia, jenis kelamin, pendapatan, perkiraan kekayaan, dan kelas sosial. Sekarang kita bisa mendapatkan informasi tentang aktivitas e-commerce, aktivitas keuangan, media sosial, minat (hobi, politik, agama), dll dari orang yang sama. Sangat mungkin bahwa atribut ini akan mempengaruhi perilaku pembelian. Oleh karena itu, menurut survei, ada lebih dari 1000 responden dengan sekitar 1000 responden dengan atribut.

Data tol adalah data yang memuat jumlah pengamatan yang banyak. Tentang Pasar Modal; Penyedia data seperti Bloomberg dan Refinitiv menyediakan akses data dalam (hampir) detik. Karena itu, hanya satu instrumen keuangan dan satu tahun. Pengamatan 31.536.000 detik dimungkinkan. Bayangkan membuat observasi panel. 10 perangkat dalam 5 tahun. Ini berarti lebih dari 160 juta pengamatan. Bagaimanapun, data besar adalah format yang paling kompleks, kombinasi dari dua di atas. Analisis efek perilaku atribut (seperti komentar Twitter) pada keputusan pembelian saham dilakukan. Dalam situasi ini, tentu saja, para peneliti sedang mengerjakan jutaan, jika bukan miliaran, pengamatan. Hasil analisis harus tersedia dalam waktu singkat. Tentu saja, spreadsheet seperti MS Excel tidak cocok untuk jenis karakter data ini. Teknik khusus yang diperlukan untuk mengambil, menyimpan, melihat, dan mempersiapkan (untuk analisis) dikenal sebagai Branch Big Data Analytics. Oleh karena itu, banyak atribut yang dapat mempengaruhi perilaku juga dapat berarti bahwa pengembangan teori deduktif menghadapi tantangan yang berat. Sekali lagi, menyimpulkan secara sistematis dan logis hubungan kausal antara variabel X dan Y adalah pemikiran abstrak. Ini membuatnya sangat sulit untuk diterapkan pada banyak properti dan metode pengembangan ilmiah. Dunia sains, sebaliknya, telah beralih ke pendekatan praktis. Jika data menunjukkan pola hubungan antar variabel, aktivitas konseptual baru dimobilisasi untuk membenarkan fenomena tersebut.

4. Simpulan

Karl Raimund Popper telah berhasil membimbing sains ke arah yang lebih kompleks dan lebih baik. Kehadirannya dalam dunia ilmu pengetahuan modern memberikan warna baru bagi perkembangan ilmu pengetahuan itu sendiri. Ide-idenya membuat sains lebih terbuka. Premis Popper mengasumsikan bahwa kritik adalah upaya terbaik untuk meningkatkan kualitas teori. Oleh karena itu, pengetahuan ilmiah harus selalu siap untuk dikritik.

Neural Network, Support Vector Machine, Linear and Logistics Regression, and Classification Trees adalah beberapa penerapan algoritma supervised learning dalam dunia *data science* bisa terjebak menjadi *pseudoscience* akibat adanya kesalahan dalam pemakaian dan pengolahan data pada penelitian. Untuk mencegah hal ini diterapkan suatu sistem demarkasi, yaitu memberlakukan sistem saintifik yang dapat dilakukan proses falsifikasi berdasarkan pengalaman sesuai premis dari Popper.

Melalui demarkasi popper, dapat diketahui bahwa suatu sistem dapat dikatakan sebagai sesuatu yang bersifat saintifik jika hal tersebut bisa dilakukan pengujian. Dalam hal ini, pengujian yang dimaksud adalah dapat diuji dengan cara apakah sesuatu tersebut mempunyai suatu kemungkinan difalsifikasi berdasarkan pengalaman. Untuk memperkuatnya dapat juga menggunakan prinsip 4 teori kebenaran yaitu koherensi, korespondensi, pragmatis dan konsensus. Berdasarkan hal ini, suatu penelitian di bidang *Data Science* dan *Machine Learning* jika tidak memperhatikan dasar-dasar pengetahuan yang tepat dan dilakukan dengan baik dapat berubah menjadi *pseudoscience*.

5. Daftar Pustaka

- Atabik, A. (2014). Teori Kebenaran Perspektif Filsafat Ilmu: Sebuah Kerangka Untuk Memahami Konstruksi Pengetahuan Agama. *Fikrah*, 2(1), 253-271.
- Dochmie, M. R. (2018). Keilmiahan Ilmu-Ilmu Islam Ditinjau dari Prinsip Falsifikasi Karl Popper. *Prosiding Konferensi Integrasi Interkoneksi Islam dan Sain*, 1, 145-150.
- Greg Yangy, J. P.-D. (2019). A Mean Field Theory of Batch Normalization. *A Conference Paper at ICLR 2019*.
- Idaman. (2019). Relasi Kuasa-Pengetahuan dalam Sistem Ketatanegaraan di Kerajaan Konawe Abad ke-XVII: Telaah Epistemologi Siwole Mbatohu. *Halu Oleo Law Review*, 132-156.
- Jimy. (2019, 8 18). *Medium*. Retrieved 11 10, 2021, from <https://bit.ly/3naHJSj>
- Krizhevsky, A. a. (2017). ImageNet Classification with Deep Convolutional Neural Networks. *Association for Computing Machinery*, 84-90.
- Kumar, S., Dhanda, N., & Pandey, A. (2018). Data Science — Cosmic Infoset Mining, Modeling and Visualization. *2018 International Conference on Computational and Characterization Techniques in Engineering & Sciences (CCTES)*, 1-4.
- Lima, T. K. (2020, 10). *KBBI Daring*. (Badan Pengembangan dan Pembinaan Bahasa) Retrieved 11 13, 2021, from <https://bit.ly/3Hj8Y5u>
- Mahayana, D. (2018). *Catatan Kuliah Filsafat Ilmu Pengetahuan*. Bandung: ITB Press.
- Martínez-Plumed, F., Contreras-Ochando, L., Ferri, C., & Hernández-Orallo, J. (2021). CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. *IEEE Transactions on Knowledge and Data Engineering*, 33(8), 3048-3061.
- Mohammad Rastegari, V. O. (2016). XNOR-Net: ImageNet Classification Using Binary Convolutional Neural Networks. *European Conference on Computer Vision*, 525-542.
- Nielson, B. a. (2021). Induction, Popper, and machine learning.
- Nur, M. (2012). Revitalisasi Epistemologi Falsifikasi. *Jurnal Agama dan Hak Azazi Manusia*, 2(1), 1-14.
- Olavsrud, T. (2020, 10 22). *5 Famous Analytics and AI Disasters*. Retrieved 11 11, 2021, from CIO.com - Tech News, Analysis, Blogs, Video: <https://bit.ly/31NT2Ys>
- Padli, M., & Mustofa, M. (2021). Kebenaran dalam Perspektif Filsafat Serta Aktualisasinya dalam Menyaring Berita. *Jurnal Filsafat Indonesia*, 4(1), 78-88.
- Rahman, S. (2017). Relevansi Epistemologi Karl R. Popper. *Komunike*, Volume IX(2), 138-149.
- Ray, S. (2019). A Quick Review of Machine Learning Algorithms. *International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, 35-39.
- Suyanto. (2018). *Machine Learning*. Bandung: Penerbit Informatika.
- Team, M. o. (2020, 11 3). *Aplikasi Machine Learning di Bidang Riset Keuangan*. (Universitas Bina Nusantara) Retrieved 11 9, 2021, from <https://bit.ly/30IYrFj>
- Warin, T. a. (2021). Machine Learning in Finance: A Metadata-Based Systematic Review of the Literature. *Journal of Risk and Financial Management*, 14(7).
- Yanghao Li, N. W. (2017). Adaptive Batch Normalization for Practical Domain Adaptation. *Pattern Recognition*.
- Yiwen Zhao, L. W. (2018). The Application of Convolution Neural Networks in Sign Language Recognition. *2018 Ninth International Conference on Intelligent Control and Information Processing (ICICIP)*, 269-272.