Analisis Perbandingan Algoritma Semi-Supervised Learning dalam Klasifikasi Adiksi Smartphone dengan Keterbatasan Data Terlabel

Triadi Widianto[1], Nani Purwati[2],Hidayat Muhammad Nur[3]

[1,2,3]Universitas Bina Sarana Informatika
e-mail: [1]triadi.trw@bsi.ac.id, [2]nani.npi@bsi.ac.id,[3]hidayat.hmm@bsi.ac.id

Artikel Info : Diterima : 02-06-2026 | Direvisi : 00-00-0000 | Disetujui : 05-06-2026

Abstrak - Penelitian ini menganalisis perbandingan kinerja algoritma *Semi-Supervised Learning* (SSL), *Support Vector Machine* (SVM), dan *Neural Network* dalam klasifikasi adiksi smartphone pada kondisi keterbata- san data terlabel. Dataset yang digunakan berjumlah 7.500 sampel dengan 15 variabel perilaku digital dan gaya hidup. Skenario eksperimen disusun secara realistis dengan hanya menggunakan 2% data latih berlabel dan 98% data tidak berlabel untuk merepresentasikan fenomena *label scarcity*. Hasil pengujian menunjukkan bahwa SSL mencapai performa terbaik dengan akurasi sebesar 88,09%, melampaui SVM (86,13%) dan *Neural Network* (85,96%). Analisis *confusion matrix* mengonfirmasi bahwa SSL mem- berikan prediksi yang lebih seimbang antarkelas serta mampu menurunkan kesalahan prediksi kritis dibanding model *supervised* murni. Meskipun demikian, terdapat *trade-off* di mana SSL membutuhkan biaya komputasi yang jauh lebih tinggi, dengan waktu pelatihan mencapai 21,76 detik berbanding SVM yang hanya 0,006 detik. Temuan ini menegaskan bahwa pemanfaatan data tidak berlabel melalui mekanisme *pseudo-labeling* merupakan strategi efektif untuk meningkatkan kualitas deteksi dini adiksi smartphone ketika proses pelabelan data oleh ahli bersifat mahal dan terbatas.

Kata Kunci : Adiksi Smartphone, *Semi-Supervised Learning*, *Label Scarcity*.

Abstracts - *This study analyzes and compares the performance of Semi-Supervised Learning (SSL), Support Vector Machine (SVM), and Neural Network algorithms in classifying smartphone addiction under conditions of limited labeled data. The dataset used comprises 7,500 samples with 15 digital behavior and lifestyle variables. Realistic experimental scenarios were designed using only 2% labeled training data and 98% unlabeled data to represent the phenomenon of label scarcity.The test results indicate that SSL achieved the best performance with an accuracy of 88.09%, outperforming SVM (86.13%) and Neural Networks (85.96%). Confusion matrix analysis confirms that SSL provides more balanced predictions across classes and is capable of reducing critical prediction errors compared to purely supervised models. Nevertheless, there is a trade-off where SSL requires significantly higher computational costs, with training times reaching 21.76 seconds compared to SVM, which only took 0.006 seconds.These findings confirm that utilizing unlabeled data through a pseudo-labeling mechanism is an effective strategy to improve the quality of early detection for smartphone addiction when the data labeling process by experts is costly and limited.*

Keywords : Smartphone Addiction, Semi-Supervised Learning, Label Scarcity.

PENDAHULUAN

Perkembangan teknologi digital telah mendorong peng-gunaan smartphone menjadi bagian integral dari aktiv-itas belajar, kerja, dan interaksi sosial. Namun, inten-sitas penggunaan yang tinggi juga meningkatkan risiko perilaku adiktif yang berdam- pak pada kesehatan mental, fungsi kognitif, kuali- tas tidur, serta performa akademik dan produk- tivitas individu. Sejumlah studi terbaru me- nunjukkan bahwa problematic smartphone use muncul pada beragam kelompok usia, mulai dari remaja hingga dewasa, dengan pola faktor risiko yang kompleks dan saling berinteraksi (Kim et al., 2024; Rochat et al., 2025; Vera Cruz et al., 2023).

Dalam konteks penelitian prediktif, machine learn-ing telah banyak digunakan untuk men- gidentifikasi faktor-faktor yang berasosiasi dengan adiksi smartphone, seperti depresi, fear of missing out (FoMO), kontrol diri, preferensi penggunaan, dan karakteristik perilaku digi-tal (Arpaci, 2023; Chen et al., 2023; Xiao et al., 2023; Zhang et al., 2025). Pendekatan ini memberikan kemam- puan pemodelan nonlinier dan interaksi fitur yang lebih baik dibanding analisis statistik konvension- al. Meski demikian, sebagian besar penelitianmasih menggunakan skema supervised learning yang mengandalkan data berlabel dalam jumlah memadai.

Permasalahan utama pada skenario nyata adalah keter-batasan data berlabel (label scarcity). Pada kasus adiksi smartphone, proses pelabelan umumnya memerlukan in-strumen psikometrik dan/atau asesmen ahli, sehingga mahal, me- makan waktu, serta sulit diskalakan. Seba-liknya, data tidak berlabel dari jejak perilaku digital re- latif melimpah. Ketimpangan ini menyebabkan model supervised seperti SVM dan neural network berpotensi mengalami penurunan generalisasi ketika proporsi data berlabel sangat kecil.

Di sisi lain, pendekatan semi-supervised learning (SSL) menawarkan solusi dengan memanfaatkan kombinasi data berlabel dan tidak berlabel me- lalui mekanisme seperti pseudo-labeling, con- sistency learning, dan rep-resentasi self- supervised. Literatur lintas domain me- nunjukkan bahwa SSL efektif meningkatkan per-forma klasifikasi pada kondisi label terbatas (Bakker et al., 2026; Saad et al., 2025; Stevanoska et al., 2025; Tang et al., 2026). Namun, bukti empiris yang secara spesifik mem-bandingkan SSL dengan model supervised klasik pada klasifikasi adiksi smartphone berbasis data tabular masih terbatas.

Berdasarkan celah tersebut, penelitian ini mengkaji per-bandingan kinerja Semi-Supervised Learning (SSL), Sup-port Vector Machine (SVM), dan Neural Network un-tuk klasifikasi adiksi smartphone pada skenario ekstrem keterbatasan label (2% data berlabel dan 98% data tidak berla- bel). Kontribusi penelitian terletak pada:

(1) evaluasi komparatif tiga pendekatan pada kondisiyang menyerupai implementasi nyata,

(2) analisis per-forma tidak hanya dari akurasi tetapi juga keseimban-gan prediksi antarkelas melalui con- fusion matrix, dan

(3) pembahasan trade-off antara peningkatan kinerja prediksi dan biaya komputasi. Hasil kajian ini diharap-kan memberi dasar metodologis bagi pengembangan sis-tem deteksi dini adiksi smartphone yang lebih adaptif dan efisien dalam kondisi keterbatasan data berlabel.

LANDASAN PUSTAKA

1. Adiksi Smartphone dan Faktor Risiko

Adiksi smartphone dipahami sebagai pola penggunaan berlebihan yang menimbulkan gangguan kontrol, keter-gantungan psikologis, serta dampak negatif pada fungsi harian. Studi terkini menunjukkan bahwa problematic smartphone use berkaitan dengan faktor psikologis (depresi, stres, psikopatologi), faktor sosial (FoMO), serta kebiasaan perilaku digital (Rochat et al., 2025; Xiao et al., 2023; Zhang et al., 2025). Pada populasi remaja dan mahasiswa, pendekatan machine learning juga telah di- gunakan untuk mengidentifikasi prediktor penting adiksi smartphone secara lebih komprehensif dibanding pen-dekatan linier tunggal (Duan et al., 2021; Li et al., 2025; Zhou, 2025).

Secara teoritis, SSL bekerja dengan asumsi bahwa struk-tur distribusi data tidak berlabel memuat in-formasi pent-ing tentang batas keputusan kelas. Dengan meman-faatkan struktur ini, model dapat mempelajari represen-tasi yang lebih baik diband- ing hanya mengandalkan data berlabel dalam jumlah kecil. Implikasi praktisnya adalah pening- katan generalisasi model pada konteks nyata yang tidak memungkinkan pelabelan masif.

2. Klasifikasi Adiksi Smartphone Berbasis Ma-chine Learning

Berbagai algoritma supervised telah diimplemen- tasikan untuk memprediksi adiksi smartphone, termasuk deci-sion tree, ensemble learning, SVM, hingga artificial neu-ral network (Arpaci, 2023; Kumar et al., 2025; Zaeni et al., 2021). SVM dikenal stabil pada data berdimensi menengah dan efektif dalam pemisahan kelas dengan margin maksimal, sedangkan neural network unggul dalam menangkap relasi nonlinier yang kompleks. Di sisi lain, beberapa studi menegaskan bahwa kualitas model sangat dipengaruhi oleh ketersediaan dan kualitas label, sehingga performa model supervised dapat menurun ketika data berlabel terbatas (Shifani et al., 2025).

3. Semi-Supervised Learning untuk Kondisi Label Scarcity

Semi-supervised learning memanfaatkan data berlabel dan tidak berlabel secara simultan untuk meningkatkan performa prediksi. Dalam konteks data tabular dan skenario low-data, be- berapa pendekatan SSL seperti autoencoder- based SSL, pseudo-labeling, dan self-supervised tabular representation menunjukkan pen-ingkatan akurasi maupun ketahanan model (Almaraz-Rivera et al., 2025; Stevanoska et al., 2025). Penelitian pada domain lain juga mem-perkuat bahwa strategi pseudo-labeling dan consistency learning mampu mengatasi

keterbatasan label secara efektif (Elsayed et al., 2026; Tang et al., 2026; Wang et al., 2025).

4. Kesenjangan Riset dan Posisi Penelitian

Walaupun riset adiksi smartphone berbasis machine learning berkembang pesat, mayoritas studi berfokus pada identifikasi faktor risiko atau pred- iksi berbasis su-pervised learning. Sementara itu, kajian yang secara langsung membandingkan SSL, SVM, dan neural net-work pada skenario label scarcity ekstrem untuk data tab-ular adiksi smartphone masih jarang dilaporkan. Oleh kare- na itu, penelitian ini menempatkan diri pada irisan dua kebutuhan: (1) kebutuhan aplikasi kesehatan digital untuk deteksi dini yang hemat biaya pelabelan, dan (2) kebutuhan metodologis untuk mengevaluasi secara adil model supervised dan semi-supervised pada kondisi op-erasional realistis.

Temuan penelitian ini diharapkan memperkaya liter-atur dengan bukti empiris bahwa pemanfaa tan data tidak berlabel dapat menjadi strategi praktis untuk meningkatkan kualitas klasifikasi adiksi smartphone tanpa ketergantungan penuh pada pelabelan manual.

METODE PENELITIAN

1. Desain Penelitian

Penelitian ini menggunakan desain komparatif kuanti-tatif untuk mengevaluasi kinerja tiga pen- dekatan klasi-fikasi, yaitu Semi-Supervised Learn- ing (SSL), Support Vector Machine (SVM), dan Neural Network, pada ka-sus klasifikasi adiksi smartphone dengan keterbatasan data berlabel. Fokus evaluasi diarahkan pada kemam-puan model dalam mempertahankan performa prediksi ketika proporsi data berlabel sangat rendah.

2. Dataset dan Variabel Penelitian

Dataset yang digunakan berjudul Smartphone Usage and Addiction Analysis diperoleh dari platform Kaggle (https://www.kaggle.com/datasets/bhadramohit/smartphone-usage-and-addiction-dataset). Dataset terdiri atas 7.500 sampel dengan 15 variabel yang merep- resentasikan pola perilaku digital dan gaya hidup pengguna smartphone. Variabel target berupa ke- las adiksi smartphone (misal-nya kategori adiktif dan non-adiktif) yang digunakan se-bagai label klasifikasi.

Secara operasional, penelitian ini menempatkan dataset sebagai data tabular terstruktur sehingga seluruh model diuji dalam kondisi input yang sa- ma. Pendekatan ini penting untuk memastikan bahwa perbandingan per-forma antarmodel bersi- fat adil dan tidak dipengaruhi perbedaan format data.

3. Tahapan Pra-pemrosesan Data

Tahapan pra-pemrosesan dilakukan untuk menjaga kual-itas data sebelum pelatihan model:

1. Pemeriksaan kualitas data, termasuk nilai hilang, duplikasi, dan outlier.
2. Transformasi fitur kategorik menjadi numerik meng-gunakan teknik encoding yang sesuai.
3. Normalisasi atau standarisasi fitur numerik agar skala fitur tidak mendominasi proses pembelajaran model.
4. Pemisahan fitur prediktor dan label target.

Seluruh tahapan pra-pemrosesan diterapkan kon- sisten pada ketiga model agar perbedaan hasil benar-benar be-rasal dari karakteristik algoritma, bukan perbedaan per-lakuan data.

4. Skenario Eksperimen Label Scarcity

Eksperimen dirancang untuk merepresentasikan kondisi realistis di mana data berlabel sulit di- peroleh. Proporsi data dibagi menjadi:

1. 2% data berlabel sebagai data latih berlabel.
2. 98% data tidak berlabel yang dimanfaatkan penuh oleh pendekatan SSL.

Pada skenario ini, SVM dan Neural Network di- latih hanya menggunakan data berlabel, se- dangkan SSL me-manfaatkan data berlabel dan tidak berlabel melalui mekanisme self-training berbasis pseudo-labeling. Den-gan demikian, evaluasi dapat menunjukkan manfaat praktis penggunaan data tidak berlabel pada kondisi ke-langkaan label.

5. Pengembangan Model

Model yang dibandingkan dalam penelitian ini adalah:

1. **Support Vector Machine (SVM)** sebagai base-line supervised learning yang kuat pada data tabu-lar.

2. **Neural Network (Multi-Layer Perceptron/MLP)** sebagai baseline nonlinier su- pervised learning.
3. **Semi-Supervised Learning (SSL) berbasis self-training** dengan alur pelatihan iteratif:
   - model awal dilatih menggunakan data berlabel,
   - model menghasilkan prediksi pada data tidak berlabel,
   - sampel dengan tingkat keyakinan tinggi diberi pseudo-label
   - data pseudo-label ditambahkan ke data latih,
   - proses diulang hingga kriteria berhenti terca-pai.

Untuk menjaga validitas perbandingan, pengaturan hy-perparameter ditentukan secara sistematis (misalnya melalui validasi silang atau pencarian grid sederhana) dengan protokol yang setara untuk seluruh model.

6. Metrik Evaluasi

Kinerja model dievaluasi menggunakan metrik klasifikasi berikut

1. **Akurasi** untuk mengukur proporsi prediksi benar secara keseluruhan.
2. **Precision** untuk menilai ketepatan prediksi kelas positif.
3. **Recall** untuk menilai kemampuan model menangkap kasus positif
4. **F1-score** sebagai rata-rata harmonik preci- sion dan recall
5. **Confusion matrix** untuk menganalisis dis- tribusi true positive, true negative, false posi- tive, dan false negative.
6. **Waktu komputasi** untuk membandingkan efisiensi pelatihan antarmodel.

Penggunaan kombinasi metrik ini memungkinkan evalu-asi yang tidak hanya berfokus pada akurasi, tetapi juga kualitas keseimbangan prediksi antarkelas

7. Prosedur Pengujian dan Analisis

Prosedur pengujian dilakukan dalam urutan beri- kut:

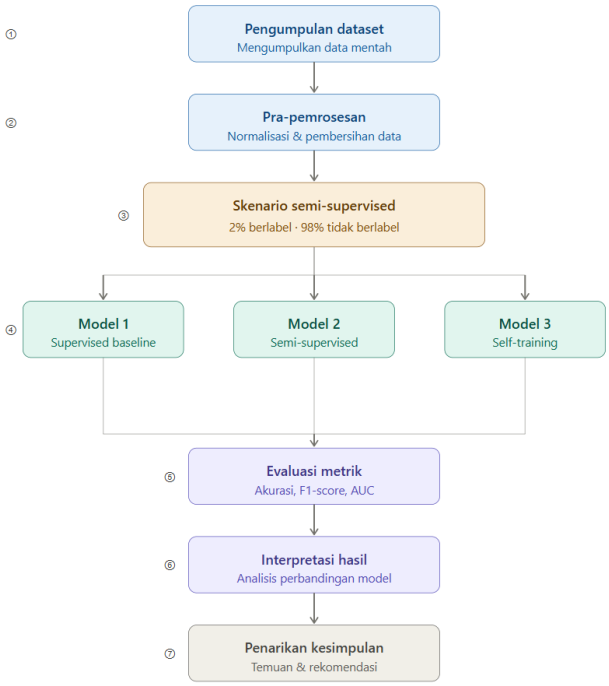
1. Menyiapkan data dan pra-pemrosesan.
2. Menerapkan skenario label scarcity (2% berlabel, 98% tidak berlabel).
3. Melatih model SVM, MLP, dan SSL sesuai protokol masing-masing.
4. Melakukan pengujian pada data uji yang sa- ma.
5. Menghitung seluruh metrik evaluasi.
6. Membandingkan hasil secara kuantitatif dan men-ganalisis confusion matrix tiap model.

Analisis akhir diarahkan untuk menjawab dua per-tanyaan utama: (1) model mana yang paling efektif pada kondisi label scarcity, dan (2) bagaimana trade-off an-tara performa prediksi dan biaya komputasi pada masing-masing pendekatan

8. Alur Kerja Penelitian

Alur ini disusun untuk memastikan replikasi ek- sperimen dapat dilakukan dengan langkah yang jelas dan konsisten.

Secara ringkas, alur kerja penelitian adalah:



Gambar 1. Alur Kerja Penelitian

HASIL PENELITIAN

1. Kinerja Utama Model

Berdasarkan hasil pengujian pada skenario data terlabel 2%, model SSL (self-training) menunjuk- kan kinerja ter-baik pada metrik agregat. Nilai akurasi tertinggi dicapai SSL sebesar 88,09%, dii- kuti SVM sebesar 86,13% dan Neural Network sebesar 85,96%. Pada metrik F1-score, SSL juga berada pada posisi tertinggi (91,78%), diband-ing SVM (90,75%) dan Neural Network (90,13%).

Pada precision, SSL (88,84%) dan Neural Net- work (88,80%) relatif setara dan lebih tinggi dibanding SVM (85,20%). Sebaliknya, pada recall SVM berada di posisi tertinggi (97,08%), disusul SSL (94,93%) dan Neural Net-work (91,50%). Temuan ini menunjukkan bahwa SVM cenderung sangat agresif dalam menangkap kelas positif, teta- pi dibayar dengan penurunan precision.

Tabel 1. Ringkasan Hasil Pengukuran Model.

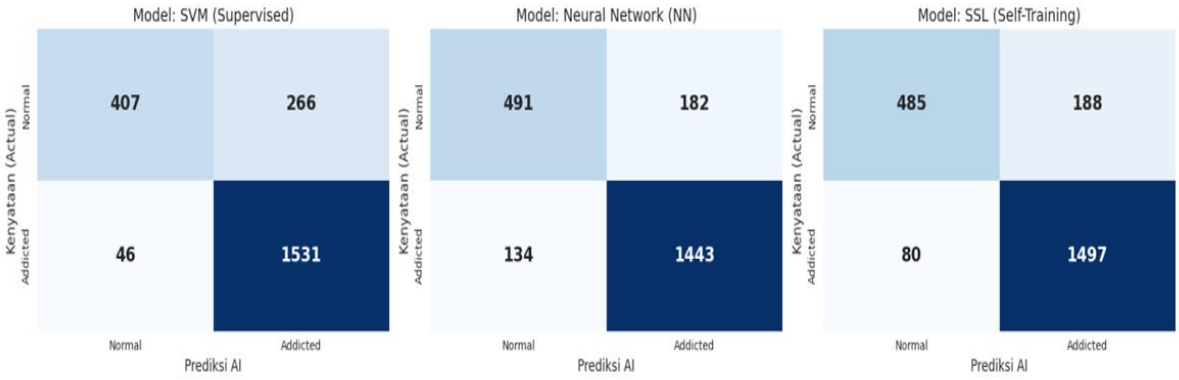
Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC	Time (s)
SVM (Supervised)	86.13%	85.20%	97.08%	90.75%	93.45%	0.006327
Neural Network (NN)	85.96%	88.80%	91.50%	90.13%	92.81%	1.443228
SSL (Self-Training)	88.09%	88.84%	94.93%	91.78%	93.29%	21.761957

2. Analisis Confusion Matrix

Hasil confusion matrix memperlihatkan pola kesalahan yang berbeda antar model:

1. **SVM**: menghasilkan false negative paling rendah (46), namun false positive tinggi (266)
2. **Neural Network**: menurunkan false positive men-jadi 182, tetapi false negative meningkat menjadi 134.
3. **SSL**: memberikan kompromi paling seimbang, den-gan false positive 188 dan false negative 80, serta jumlah prediksi benar tertinggi (1982 dari 2250 data uji).

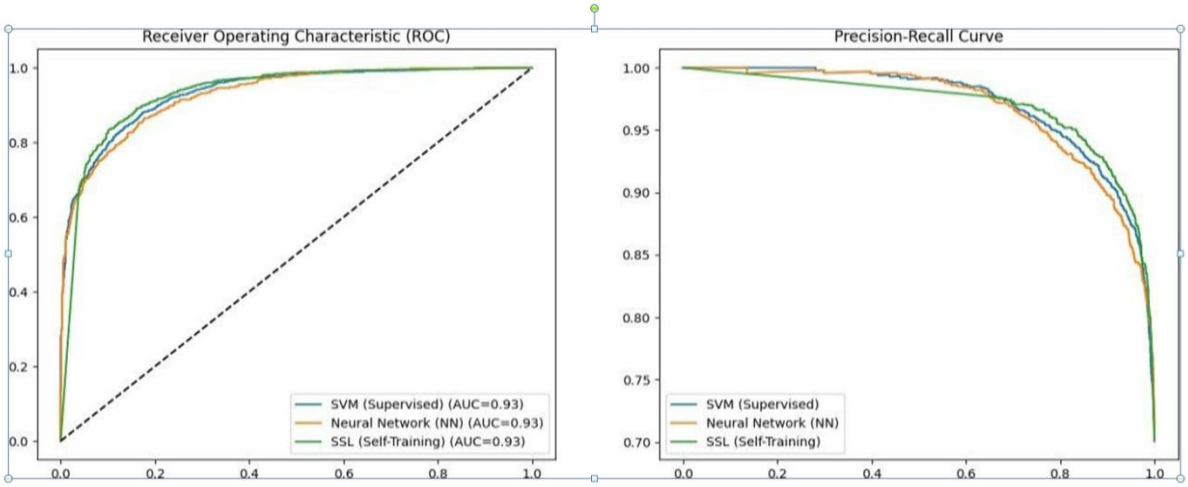
Secara praktis, SSL menunjukkan kemampuan



Gambar 2. Perbandingan Confusion Maatrix (Data Latih 2%)

3. ROC-AUC dan Precision-Recall Curve

Kurva ROC menunjukkan ketiga model memiliki ke-mampuan diskriminasi yang sangat baik dan relatif berdekatan (sekitar AUC 0,93). Namun, dari kurva precision-recall terlihat bahwa SSL cenderung mempertahankan precision yang lebih stabil pada rentang recall menengah tinggi dibanding dua model supervised. Pola ini konsis- ten dengan capaian F1-score SSL yang tertnggi.



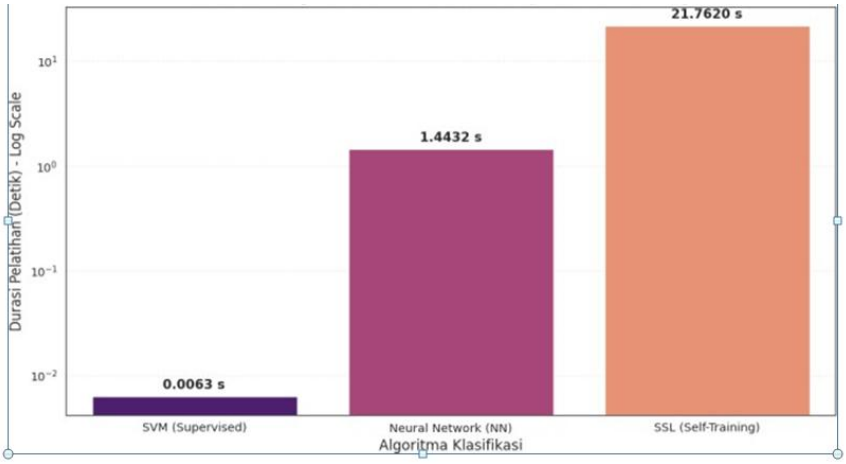
Gambar 3. Kurva ROC dan precision-recall untuk ketiga model.

4. Efisiensi Waktu Komputasi

Perbandingan waktu pelatihan menunjukkan perbedaan yang sangat signifikan:

1. SVM: 0,0063 detik.
2. Neural Network: 1,4432 detik.
3. SSL: 21,7620 detik.

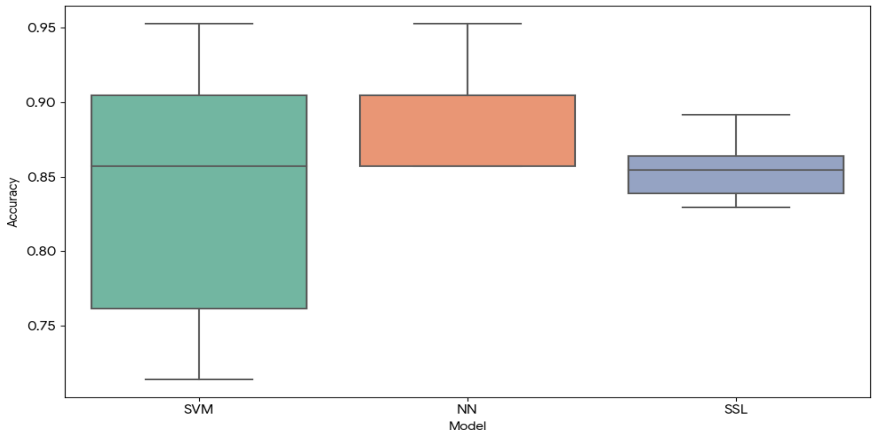
Hasil ini menegaskan adanya trade-off antara performa prediksi dan biaya komputasi. SSL memberikan akurasi tertinggi, tetapi membutu kan durasi pelatihan paling lama karena proses iteratif pseudo-labeling.



Gambar 4. Perbandingan efisiensi waktu komputasi antar model (skala log).

5. Stabilitas Akurasi (5-Fold Cross- Validation)

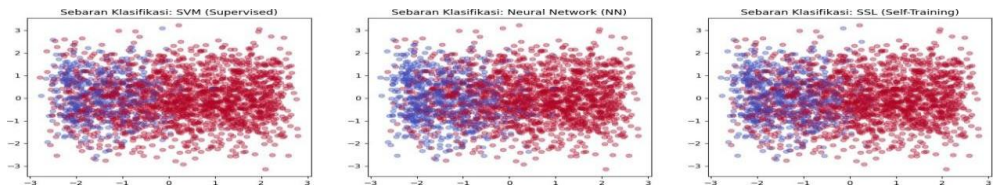
Boxplot menunjukkan SSL memiliki rentang akurasi antar-fold yang paling sempit, menandakan variabilitas performa yang lebih rendah. SVM memiliki rentang paling lebar, mengindi- kasikan sensitivitas lebih tinggi terhadap variasi pembagian data. Neural Network be-rada di ten- gah, dengan variasi yang lebih baik dari SVM tetapi masih kurang stabil dibanding SSL



Gambar 5. Stabilitas akurasi model pada 5-fold cross-validation.

6. Pola Data dan Sebaran Klasifikasi

Visualisasi sebaran klasifikasi memperlihatkan adanya area tumpang tindih antarkelas normal dan adiktif. Kon-disi ini menegaskan bahwa masalah klasifikasi tidak trivial. Pada kondisi seperti ini, pemanfaatan data tidak berlabel oleh SSL membantu model membentuk batas keputusan yang lebih adaptif dibanding pendekatan supervised murni.



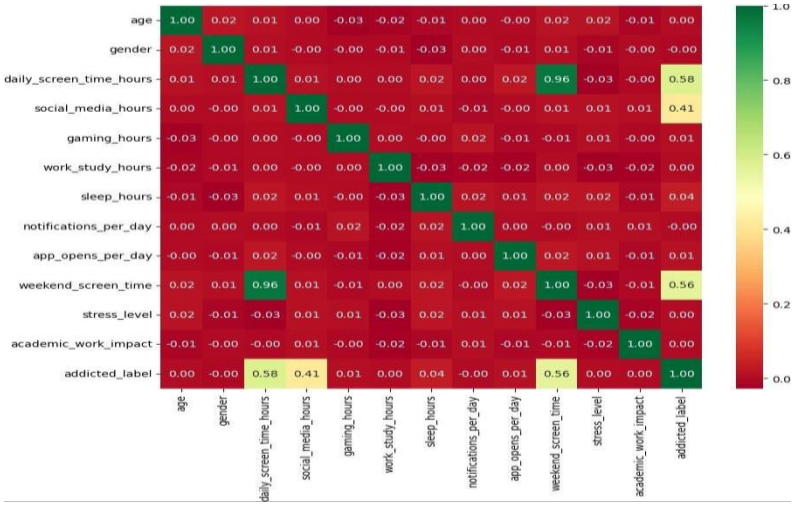
Gambar 6. Sebaran klasifikasi pada SVM, Neural Network, dan SSL

7. Analisis Hubungan Variabel dan Pent- ingnya Fitur

Heatmap korelasi menunjukkan bahwa hubungan paling kuat terhadap label adiksi muncul pada variable terkait durasi penggunaan layar, yaitu *daily screen time hours* (0,58) dan *weekend screen time* (0,56), diikuti *social media hours* (0,41). Hasil ini mengindikasikan bahwa indikator intensitas penggunaan

smartphone tetap menjadi penjelas utama status adiksi.

Di sisi lain, analisis penghapusan fitur pada model SSL memperlihatkan bahwa akurasi paling banyak turun ketika social_media_hours di- hilangkan, lalu weekend_screen_time dan dai-ly_screen_time_hours. Konsistensi antara korelasi dan feature ablation mem-perkuat validitas sub- stantif bahwa ketiga variabel terse-but merupakan faktor dominan dalam klasifikasi adiksi smartphone.



Gambar 7. Heatmap korelasi antar variabel pada dataset penelitian.

8. Tabel Keputusan Multi-Kriteria

Ringkasan keputusan multi-kriteria dari dokumen utama penelitian adalah sebagai berikut:

1. Aspek akurasi dan F1-score: pemenang objek-tif adalah SSL (Semi-Supervised) kare- na meman-faatkan struktur data tak terlabel.
2. Aspek waktu komputasi: pemenang objektif adalah SVM (Supervised) karena proses pelatihan langsung (single-pass).
3. Aspek ketangguhan (robustness): pemenang objek-tif adalah SSL (Semi-Supervised) kare-na varians akurasi paling rendah pada cross- validation.
4. Aspek kemudahan implementasi: pemenang objektif adalah SVM/NN karena tidak me-merlukan pengat-uran threshold probabilitas.

Tabel 2. Tabel Ringkasan Perbandingan Performa

Aspek Penilaian	Pemenang Objektif	Alasan Teknis
Akurasi & F1-Score	SSL (Semi-Supervised)	Memanfaatkan struktur data tak terlabel.
Waktu Komputasi	SVM (Supervised)	Proses pelatihan langsung (*single-pass*).
Ketangguhan (Robustness)	SSL (Semi-Supervised)	Varians akurasi paling rendah pada Cross-Validation.
Kemudahan Implementasi	SVM / NN	Tidak memerlukan pengaturan *threshold* probabilitas.

PEMBAHASAN

1. Mengapa SSL Unggul pada Skenario Label Scarcity

Keunggulan SSL pada akurasi dan F1-score menun-jukkan bahwa informasi dari data tidak berlabel berhasil dimanfaatkan untuk mem- perkaya struktur keputusan model. Dengan han- ya 2% data berlabel, model super-vised cenderung mengalami keterbatasan informasi un-tuk memetakan batas kelas secara optimal. SSL mengatasi keterbatasan ini melalui pembelajaran iteratif dari pseudo-label, sehingga representasi kelas menjadi lebih kuat.

Temuan ini sejalan dengan literatur SSL pada data tabular dan skenario low-label yang menunjukkan bahwa data tidak berlabel dapat meningkatkan generalisasi ketika kualitas pseudo-label terjaga.

2. Interpretasi Trade-Off Kinerja

Jika prioritas penelitian adalah kecepatan kom- putasi, SVM menjadi kandidat paling efisien ka- rena waktu pelatihan sangat rendah. Namun, jika prioritas utama adalah kualitas prediksi kese- luruhan, terutama keseim-bangan precision-recall yang tercermin pada F1-score, SSL lebih unggul. Neural Network pada studi ini

belum mampu melampaui SSL maupun SVM secara konsisten, terutama pada skenario label sangat terbatas.

Dengan demikian, pemilihan model perlu me- nyesuaikan konteks implementasi:

1. **SVM** untuk kebutuhan inferensi/pelatihan sangat cepat dengan toleransi false positive lebih tinggi.
2. **SSL** untuk kebutuhan deteksi yang lebih aku- rat dan stabil pada kondisi keterbatasan label.

3. Hubungan Antarvariabel dan Implikasi Substantif

Heatmap korelasi menunjukkan variable yang paling terkait dengan label adiksi adalah dayly screen time hours (0,58), week end screen time (0,56), dan social_media_hours (0,41). Variabel lain cen- derung memiliki korelasi linear rendah terhadap label, yang mengindikasikan pentingnya model yang mampu menangkap relasi nonlinier dan interaksi antarfitur.

Analisis penghapusan fitur pada model SSL juga konsisten: penurunan akurasi terbesar ter- jadi saat social_media_hours dihapus, kemudian weekend_screen_time dan daily_screen_time_hours. Artinya, ketiga fitur tersebut merupakan pen- dorong utama kinerja model dalam membedakan pengguna adiktif dan non-adiktif.

4. Implikasi Praktis

Hasil penelitian memperkuat bahwa sistem de- teksi dini adiksi smartphone dapat ditingkatkan melalui pen-dekatan semi-supervised tanpa ketergantungan penuh pada pelabelan manual berskala besar. Pada implemen-tasi nyata di insti- tusi pendidikan atau layanan kesehatan digital, strategi ini relevan karena biaya pelabelan sering menjadi hambatan utama.

Meskipun demikian, peningkatan akurasi melalui SSL perlu diimbangi dengan perencanaan sumber daya kom putasi, terutama jika sistem akan dijal- ankan pada skala data yang lebih besar.

5. Keterbatasan Pembahasan

Interpretasi hasil masih berbasis satu skenario utama (2% label) dan satu konfigurasi eksperimen. Untuk memperkuat generalisasi temuan, studi lanjutan disarankan menguji beberapa rasio label tambahan, teknik SSL al-ternatif, serta evaluasi pada dataset eksternal.

6. Sintesis Keputusan Model

Jika hasil kuantitatif dan tabel keputusan dipadukan, maka strategi pemilihan model dapat dirumuskan secara kontekstual. Untuk tujuan akademik yang menekankan kualitas prediksi pa da label scarcity, SSL menjadi rekomendasi uta- ma. Untuk implementasi cepat den-gan sumber daya komputasi terbatas, SVM lebih rasional. Dengan kata lain, keputusan model pada studi ini bersifat multi-objektif: akurasi dan robustness mengarah ke SSL, sedangkan efisiensi implemen- tasi mengarah ke SVM/NN.

7. Kesimpulan

Penelitian ini menunjukkan bahwa pada kondisi data terlabel yang sangat terbatas (2%), pendeka- tan Semi-Supervised Learning (SSL) berbasis self- training mem-berikan kinerja terbaik dibanding SVM dan Neural Net-work. SSL mencapai akurasi 88,09% dan F1-score 91,78%, sekaligus menun- jukkan kestabilan performa yang lebih baik pada evaluasi lintas-fold. Hasil confu-sion matrix juga menegaskan bahwa SSL mampu men- jaga kese- imbangan kesalahan prediksi antarkelas secara lebih baik dibanding model supervised murni.

SVM tetap unggul pada efisiensi komputasi, na- mun per-forma keseluruhan masih berada di bawah SSL untuk skenario label scarcity. Dengan demikian, pemanfaatan data tidak berlabel ter- bukti menjadi strategi efektif un-tuk meningkat- kan kualitas klasifikasi adiksi smartphone ketika proses pelabelan data mahal dan terbatas.

Saran

1. Penelitian lanjutan disarankan menguji beberapa rasio data berlabel (misalnya 1%, 5%, 10%, dan 20%) untuk memetakan titik tran- sisi keunggulan SSL terhadap model super- vised.
2. Perlu dilakukan perbandingan dengan varian SSL lain, seperti consistency regularization dan graph-based SSL, agar diperoleh konfig- urasi model yang lebih optimal.
3. Studi berikutnya sebaiknya menambahkan evaluasi fairness dan interpretabilitas model untuk memas- tikan hasil prediksi dapat digunakan secara etis pada konteks pendidikan dan kesehatan digital.
4. Validasi eksternal pada dataset berbeda di- perlukan agar generalisasi model lebih kuat sebelum imple-mentasi skala nyata.

5. Untuk implementasi operasional, pemilihan model sebaiknya berbasis kebutuhan: SSL untuk akurasi dan robustness, SVM untuk ke- cepatan dan efisiensi komputasi.

REFERENSI

Almaraz-Rivera, J. G., Cantoral-Ceballos, J. A., Botero, J. F., Munoz, F. J., & Martinez, B. D. (2025). Hyphatia: A Card-Not-Present Fraud Detection System Based on Self-Supervised Tabular Learning. *IEEE Open Journal of the Computer Society*.

Arpaci, I. (2023). Predicting Problematic Smartphone Use Based on Early Maladaptive Schemas by Using Machine Learning Classification Algorithms. *Journal of Rational-Emotive and Cognitive-Behavior Therapy*.

Bakker, S., Ma, Y., & Ziabari, S. S. M. (2026). Addressing Label Scarcity: Hybrid Anomaly Detection in Mental Healthcare Billing. *Lecture Notes in Computer Science*.

Chen, J., Zhang, R., Chen, J., Hu, C., & Mao, Y. (2023). Open-Set Semi-Supervised Text Classification with Latent Outlier Softening. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 226–236). https://doi.org/10.1145/3580305.3599456

Duan, L., He, J., Li, M., Dai, J., Zhou, Y., Lai, F., & Zhu, G. (2021). Based on a Decision Tree Model for Exploring the Risk Factors of Smartphone Addiction Among Children and Adolescents in China During the COVID-19 Pandemic. *Frontiers in Psychiatry*.

Elsayed, N., Liu, J., Li, C., Diakite, A., Song, D., Osman, Y. B. M., & Wang, S. (2026). Boundary-Aware and Discrepancy-Guided Dynamic Pseudo-Labeling with Consistency Learning for Semi-Supervised 3D TOF-MRA Cerebrovascular Segmentation. *Physics in Medicine and Biology*.

Kim, K., Yoon, Y., & Shin, S. (2024). Explainable Prediction of Problematic Smartphone Use Among South Korea's Children and Adolescents Using a Machine Learning Approach. *International Journal of Medical Informatics*.

Kumar, A. K., Sailatha, K., Muskan, H. T., Krishna, G. V., & Rao, P. Y. (2025). Prediction of Smartphone Addiction Using Ensemble Algorithm. In *Synergies in Smart and Virtual Systems Using Computational Intelligence*.

Li, X., Zhang, Y., & Wang, L. (2025). Short Text Classification Based on Enhanced Word Embedding and Hybrid Neural Networks. *Applied Sciences*, *15*(9), 5102.

Rochat, L., Cruz, G. V., Aboujaoude, E., Courtois, R., Brahim, F. B., Khan, R., & Khazaal, Y. (2025). Problematic Smartphone Use in a Representative Sample of US Adults: Prevalence and Predictors. *Addictive Behaviors*.

Saad, E., Besbes, M., Zolghadri, M., Baron, C., Czmil, V., & Bourgeois, V. (2025). Enhancing Obsolescence Forecasting with Deep Generative Data Augmentation: A Semi-Supervised Framework for Low-Data Industrial Applications. *Journal of Intelligent Manufacturing*.

Stevanoska, S., Levatic, J., & Dzeroski, S. (2025). Semi-Supervised Learning from Tabular Data with Autoencoders: When Does It Work? *Machine Learning*.

Tang, M., Yao, L., Zhu, Z., Shen, B., Zeng, J., & Song, Z. (2026). Semi-Supervised Spatio-Temporal Graph Variational Autoencoder for Enhanced Industrial Process Fault Diagnosis Under Label Scarcity. *Control Engineering Practice*.

Vera Cruz, G., Aboujaoude, E., Khan, R., Rochat, L., Ben Brahim, F., Courtois, R., & Khazaal, Y. (2023). Smartphone Apps for Mental Health and Wellbeing: A Usage Survey and Machine Learning Analysis of Psychological and Behavioral Predictors. *Digital Health*.

Wang, W., Yu, C., Zhang, T., Chen, F., Liu, Y., Liu, Y., & Wu, Z. (2025). Oversized ore segmentation using SAM-enhanced U-Net with self-supervised pre-training and semi-supervised self-training. In *Expert Systems with Applications* (Vol. 285). Elsevier Ltd. https://doi.org/10.1016/j.eswa.2025.127980

Xiao, B., Parent, N., Rahal, L., & Shapka, J. (2023). Using Machine Learning to Explore the Risk Factors of Problematic Smartphone Use Among Canadian Adolescents During COVID-19: The Important Role of Fear of Missing Out (FoMO). *Applied Sciences*.

Zaeni, I. A. E., Anzani, D. R., Sudjiwanati, Kristianty, E. P., & Sheng, A. Q. (2021). Classification of the Smartphone Addiction Using the Artificial Neural Network. *2021 International Research Symposium on Advanced Engineering and Vocational Education (IRSAEVE 2021)*.

Zhang, B., Zhang, X., Zhou, Z., Liu, Y., Li, Y., & Huang, F. (2025). Moment matching of joint distributions for unsupervised domain adaptation. *Information Processing & Management*, *62*(1), 103944. https://doi.org/10.1016/j.ipm.2024.103944

Zhou, Y. (2025). Hybrid Recommendation Framework for Personalized Library Services Based on Deep Learning. In *Proceedings—2025 2nd International Conference on Intelligent Computing and Robotics, ICICR 2025* (pp. 934–938). https://doi.org/10.1109/ICICR65456.2025.00165