

Analisis Sentiment Ulasan Aplikasi Gopay Pada Google Play Store Menggunakan Algoritma Support Vector Machine

Shafa Nadhifah¹, Fadhila Nur Aini², Hadwitya Handayani Kusumawardhani³, Muhammad Yusuf Febrianto⁴

^{1,2} Informatika, Fakultas Teknik dan Ilmu Komputer Universitas Muhammadiyah Pekajangan Pekalongan

^{3,4} Manajemen Informatika, Fakultas Teknik dan Ilmu Komputer Universitas Muhammadiyah Pekajangan Pekalongan
shafanadifa111@gmail.com¹, nurainii.fadhila@gmail.com², hadwitya.hk@gmail.com³, mysuffebrianto@umpp.ac.id⁴

Abstract

Gopay is a platform application on the Google Play Store. The Gopay application has been downloaded 10 million times but only has a rating of 4.4. On the Google Play Store, application ratings are given followed by user reviews of the application. There are quite a lot of reviews and are unstructured and contain opinions from users about their satisfaction with the application, so they are often taken into consideration by potential users when choosing which application to use. Based on this, sentiment analysis was carried out using the Support Vector Machine algorithm to find out how users feel about the GOPAY application on the Google Play Store. This research uses the Knowledge Discovery in Database (KDD) method. The data used is 29,554 reviews of the GOPAY application, consisting of 23,961 positive reviews and 5,593 negative reviews. The best results were obtained in scenario 1 (90:10) using the RBF (Radial Basis Function) kernel which produced an accuracy of 65%, precision of 71.43%, recall of 50%, and f1-score of 58.82%.

Keywords : Analisis Sentimen, Support Vector Machine, Gopay

Abstraksi

GOPAY adalah salah satu aplikasi di google play store. Aplikasi Gopay sudah diunduh sebanyak 10 juta kali namun hanya memiliki rating 4,4. Pada google play store pemberian rating aplikasi diikuti dengan ulasan dari para pengguna terhadap aplikasi. Ulasan yang ada cukup banyak dan tidak terstruktur serta mengandung opini dari para pengguna tentang kepuasan mereka terhadap aplikasi tersebut sehingga sering dijadikan pertimbangan oleh calon pengguna untuk memilih aplikasi yang digunakan. Berdasarkan hal tersebut, maka dilakukan analisis sentimen menggunakan algoritma *Support Vector Machine* untuk mengetahui bagaimana sentimen para pengguna terhadap aplikasi GOPAY di google play store. Penelitian ini menggunakan metode Knowledge Discovery in Database (KDD). Data yang digunakan merupakan ulasan aplikasi GOPAY sebanyak 29554 data yang terdiri dari 23961 ulasan positif dan 5593 ulasan negatif. Hasil terbaik diperoleh pada skenario 1 (90:10) menggunakan kernel RBF (Radial Basis Function) yang menghasilkan accuracy 65% precision 71,43%, recall 50%, dan f1-score 58,82%.

Keywords : Analisis Sentimen, Support Vector Machine, Gopay

1. PENDAHULUAN

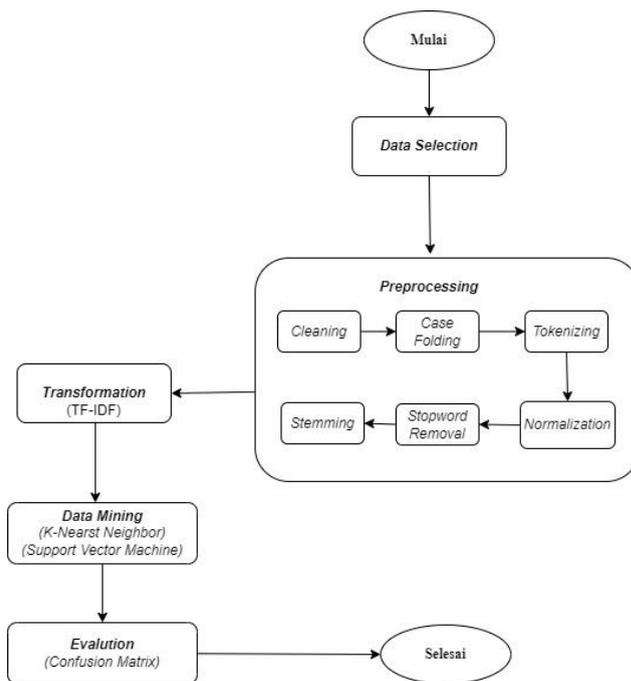
Ulasan aplikasi di Google Play Store memiliki banyak dampak positif bagi konsumen yang hendak menggunakan suatu produk. Berikut adalah beberapa pentingnya ulasan aplikasi di Google Play Store Informasi Pengalaman Pengguna: Ulasan memberikan informasi mengenai pengalaman pengguna yang sebenarnya dengan menggunakan aplikasi tersebut. Ini dapat membantu calon pengguna untuk memahami kelebihan dan kekurangan dari sudut pandang pengguna lain. Keandalan dan Kualitas: Ulasan dapat memberikan gambaran tentang keandalan dan kualitas aplikasi. Jika banyak ulasan positif, ini dapat memberikan keyakinan kepada konsumen bahwa aplikasi tersebut handal dan berkualitas. Saran Perbaikan: Ulasan juga seringkali berisi saran perbaikan dari pengguna yang dapat membantu pengembang untuk terus memperbaiki dan meningkatkan kualitas aplikasi. Hal ini menciptakan siklus umpan balik yang positif. Keamanan dan Privasi: Ulasan dapat mencakup aspek keamanan dan privasi dari aplikasi. Ini adalah kekhawatiran penting bagi banyak pengguna, dan melihat ulasan dapat membantu pengguna untuk membuat keputusan yang lebih informasional. Pembaruan dan Perbaikan Bug: Pengembang sering merespons langsung pada ulasan untuk memberikan pembaruan atau memperbaiki bug yang dilaporkan oleh pengguna. Ini menciptakan komunikasi antara pengembang dan pengguna

yang sangat bermanfaat. Pilihan Aplikasi Alternatif: Ulasan juga bisa memberikan perbandingan dengan aplikasi serupa. Hal ini membantu konsumen untuk memilih aplikasi yang paling sesuai dengan kebutuhan dan preferensi mereka. Kredibilitas Pengembang: Melihat ulasan dapat memberikan gambaran tentang kredibilitas pengembang. Jika pengembang terus menerima ulasan positif dan memberikan tanggapan yang baik terhadap masukan pengguna, ini dapat meningkatkan kepercayaan konsumen. Keputusan Pembelian: Bagi konsumen yang mempertimbangkan untuk membeli versi berbayar atau melakukan pembelian dalam aplikasi, ulasan dapat menjadi faktor penting dalam pengambilan keputusan. Komunitas Pengguna: Ulasan juga dapat membantu menciptakan komunitas pengguna di sekitar aplikasi. Pengguna dapat berbagi tips, trik, atau solusi masalah melalui ulasan, menciptakan atmosfer kolaboratif. Mengurangi Risiko Kekecewaan: Dengan membaca ulasan sebelum mengunduh atau membeli aplikasi, pengguna dapat mengurangi risiko kekecewaan dan merasa lebih percaya diri dengan keputusan mereka. Dengan demikian, ulasan aplikasi di Google Play Store memiliki peran yang signifikan dalam membentuk persepsi dan keputusan konsumen mengenai suatu produk atau layanan

Metode *Support Vector Machine* (SVM) adalah teknik yang umum digunakan dalam penambangan data teks untuk berbagai tujuan analisis dan klasifikasi. SVM dapat digunakan untuk tugas klasifikasi teks, seperti kategorisasi dokumen, analisis sentimen, dan pengklasifikasian teks lainnya. SVM dikenal memiliki kemampuan klasifikasi yang tinggi, terutama ketika jumlah fitur (kata-kata atau istilah) sangat besar. Data teks sering kali memiliki dimensi yang tinggi karena setiap kata dapat dianggap sebagai fitur. SVM efektif dalam menangani masalah ruang dimensi tinggi, dan dapat memberikan hasil yang baik bahkan dengan jumlah fitur yang besar. SVM cenderung lebih tahan terhadap overfitting dibandingkan dengan beberapa metode klasifikasi lainnya. Ini berguna ketika kita memiliki dataset teks yang relatif kecil atau ketika data memiliki karakteristik yang kompleks. SVM mendukung penggunaan kernel untuk mentransformasikan data ke dalam ruang fitur yang lebih tinggi. Ini memungkinkan SVM menangani data yang tidak linear dengan baik. Kernel seperti kernel linear, polinomial, dan radial basis function (RBF) dapat digunakan untuk menyesuaikan dengan bentuk data yang berbeda. SVM memiliki keunggulan dalam mengelola dataset yang besar dan kompleks. Beberapa implementasi SVM, terutama dalam konteks penambangan data teks, dapat dioptimalkan untuk meningkatkan efisiensi dalam pemrosesan data besar. SVM menghasilkan model yang cukup mudah diinterpretasi, terutama dalam konteks klasifikasi teks. Vektor dukungan (*support vectors*) yang digunakan dalam proses pembelajaran dapat memberikan wawasan tentang fitur-fitur yang paling penting dalam klasifikasi. SVM dapat diatur untuk menangani dataset yang tidak seimbang, di mana kelas-kelas yang berbeda memiliki jumlah instans yang berbeda. Ini berguna dalam banyak kasus penambangan data teks di mana kelas-kelas mungkin tidak memiliki distribusi yang merata. SVM dapat digunakan efektif dengan representasi fitur langsung (seperti representasi one-hot) atau dengan representasi sparse yang umum dalam data teks.

2. LANDASAN TEORI

Pemberian rating aplikasi di Google Play Store diikuti dengan ulasan dari para pengguna terhadap aplikasi tersebut. Ulasan tersebut mengandung opini dari para pengguna mengenai aplikasi tersebut dan calon pengguna melihat ulasan dari sebuah aplikasi sebagai pertimbangan sebelum memutuskan untuk menggunakan aplikasi tersebut (Saputra et al., 2019). Ulasan di Google Play Store ini cukup banyak dan tidak terstruktur sehingga diperlukan suatu teknik untuk mengetahui bagaimana ulasan pengguna terhadap aplikasi tersebut (Fransiska & Irham Gufroni, 2020). Oleh karena itu, diperlukan analisis sentimen pada data ulasan. Analisis sentimen adalah teknik untuk mengolah data tekstual untuk memperoleh informasi pada teks (Wahyudi & Kusumawardana, 2021). Terdapat beberapa algoritma yang dapat diterapkan dalam analisis sentimen seperti K-Nearest Neighbor (KNN), Naïve Bayes, dan



Gambar 1 Alur Penelitian

Support Vector Machine (SVM). Berdasarkan beberapa referensi, SVM menghasilkan akurasi terbaik dalam melakukan analisis sentimen dibandingkan algoritma klasifikasi yang lain. Pada penelitian sebelumnya SVM memiliki tingkat akurasi yang lebih tinggi sebesar 81,46% dibandingkan Naïve Bayes sebesar 75,41% saat diterapkan untuk analisis sentimen ulasan Bahasa Indonesia di Google Play Store (Ilmawan & Mude, 2020). Pada penelitian lain yang membandingkan SVM, Naïve Bayes, dan KNN untuk analisis sentimen gadget, SVM menghasilkan rata-rata akurasi terbaik sebesar 96,43% (Iskandar & Nataliani, 2021). SVM mempunyai fungsi yang bisa mentransformasikan data ke ruang dimensi yang lebih tinggi yaitu ruang kernel yang disebut dengan fungsi kernel trick sehingga data dapat dipisahkan dengan lebih baik dibandingkan algoritma klasifikasi lainnya (Mukarramah et al., 2021). Beberapa kernel pada SVM yaitu kernel Linear, Radial Basis Function (RBF), Sigmoid, dan Polynomial (Zuriel & Fahrurrozi, 2021). Pada penelitian analisis sentimen ulasan aplikasi Ruangguru dengan membandingkan 3 kernel algoritma SVM diperoleh hasil bahwa kernel linear merupakan kernel terbaik dengan akurasi hingga 89,7% (Irfani et al., 2020). Sedangkan dalam penelitian lainnya akurasi terbaik diperoleh oleh kernel RBF dalam analisis sentimen pada maskapai penerbangan sebesar 84,37% (Husada & Paramita, 2021). Berdasarkan penjelasan di atas, karena belum ada hasil pasti mengenai akurasi kernel terbaik dalam melakukan analisis sentimen, maka pada penelitian ini akan dilakukan perbandingan akurasi antara kernel Linear, kernel RBF, kernel Sigmoid, dan kernel Polynomial pada algoritma SVM dalam melakukan analisis sentimen ulasan aplikasi DANA untuk mengetahui bagaimana sentimen para pengguna terhadap aplikasi DANA di Google Play Store dan akurasi kernel terbaik.

3. METODOLOGI PENELITIAN

Penelitian ini menggunakan metode KDD (Knowledge Discovery in Database) karena memiliki keunggulan proses pengidentifikasian pola yang terorganisir dari sekumpulan data yang kompleks, sehingga datanya mudah untuk dipahami (Ramos et al., 2021; Rizki et al., 2021). Gambar 1 menunjukkan alur penelitian yang menerapkan metode KDD.

Alur penelitian pada gambar 1 dijelaskan sebagai berikut:

1. Data Selection

Data selection merupakan tahapan pertama dalam KDD, pada tahap ini akan dilakukan pengumpulan, seleksi, dan pelabelan data. Data dikumpulkan dari situs Google Play Store pada tanggal 26 Februari 2020 sampai 4 Desember 2023.

2. Preprocessing

Pada tahap preprocessing ini, data mentah yang telah terkumpul akan diolah menjadi data yang dapat digunakan pada tahap selanjutnya. Tahap preprocessing terdiri dari 6 proses yaitu:

- Cleaning*, merupakan proses untuk menghilangkan atribut yang tidak berpengaruh terhadap klasifikasi yaitu tanda baca, karakter kosong, dan emoji.
- Case Folding*, merupakan proses untuk mengubah semua huruf yang ada pada dokumen menjadi huruf kecil.
- Tokenizing*, merupakan proses untuk memisahkan kalimat menjadi beberapa bagian kata.
- Normalization*, merupakan proses untuk memperbaiki kesalahan yang ada pada kata seperti ejaan yang salah agar kata yang memiliki makna sama menjadi setara.
- Stopword Removal*, merupakan proses untuk menghilangkan katakata yang dianggap tidak berpengaruh terhadap kalimat.
- Stemming*, merupakan proses untuk mengubah katakata yang ada menjadi bentuk kata dasar.

3. Transformation

Tahap transformation adalah tahap mengubah data menjadi bentuk yang dapat diolah pada tahapan data mining. Pada tahap ini akan dilakukan pembagian data ke dalam 3 skenario terlebih dahulu yaitu skenario 1 (90% data training dan 10% data testing), skenario 2 (80% data training dan 10% data testing), skenario 3 (90% data training dan 10% data testing). Setelah itu akan dilakukan pembobotan kata dengan TF-IDF yang berguna untuk mengubah data berupa teks menjadi vektor bobot. Perhitungan metode TFIDF dinyatakan dalam persamaan sebagai berikut.

$$W_t = TF_{t,d} \times IDF_t = TF_{t,d} \times \log \frac{N}{DF_t}$$

4. Data Mining

Proses data mining dalam klasifikasi sentimen melibatkan beberapa tahap yang dirancang untuk mengekstraksi informasi dan pola dari data teks atau data sentimen yang relevan. Proses klasifikasi sentimen menggunakan empat kernel algoritma Support Vector, diantaranya yaitu Kernel Linear, Kernel RBF, kernel

Sigmoid dan Kernel Polymonial. Berikut merupakan persamaan dalam proses data mining menggunakan algoritma Support Vector:

- Kernel Linear

$$K(x_i x) = x_i x$$

- Kernel RBF

$$K(x_i x) = \exp\left(-\frac{\|x_i x\|^2}{2\sigma^2}\right)$$

- Kernel Sigmoid

$$K(x_i x) = (x_i x)^d$$

- Kernel Polynomial

$$K(x_i x) = \tanh(\sigma(x_i x) + c)$$

- Kernel Linear

$$K(x_i x) = x_i x$$

- Kernel RBF

$$K(x_i x) = \exp\left(-\frac{\|x_i x\|^2}{2\sigma^2}\right)$$

- Kernel Sigmoid

$$K(x_i x) = (x_i x)^d$$

- Kernel Polynomial

$$K(x_i x) = \tanh(\sigma(x_i x) + c)$$

4. HASIL DAN PEMBAHASAN

1. Data Selection

Pengumpulan data ulasan aplikasi GOPAY yang dilakukan dengan teknik scraping pada tanggal 26 Februari 2020 sampai 4 Desember 2023. menggunakan library google play scraper menghasilkan sebanyak 29554 data. Data yang telah terkumpul memiliki 5 atribut yaitu reviewId, userName, at, content, dan score. Selanjutnya dilakukan penyeleksian atribut yang digunakan yaitu hanya content. Kemudian dilakukan pelabelan data ulasan secara manual dan divalidasi oleh ahli Bahasa dan Sastra Indonesia. Pelabelan data tersebut menghasilkan 23961 ulasan positif dan 5593 ulasan negatif. Hasil pada tahap data selection dapat dilihat pada gambar 2.

	content	score	sentimen
0	Ga jelas...kok antri... aneh...kalo blm siap ...	1	0
1	seperti aplikasi penipuan hati hati ini . sang...	1	0
2	Tidak bisa di buka aplikasinya parah bnget gopay	1	0
3	Ribet asw mau buka akun aja susah baru kli ini...	1	0
4	Kok gk ada ulasan lain?	5	1

Gambar 2 Hasil Data Selection

2. Preprocessing

Pada langkah ini, data diubah agar sesuai dan dapat diolah pada langkah berikutnya. Pada tabel 1 menunjukkan hasil dari enam proses yang telah dilakukan pada tahap preprocessing.

Table 1. Contoh Hasil Preprocessing

Proses	Hasil
Data Ulasan	Tidak bisa di buka aplikasinya parah bnget gopay
Cleaning	Tidak bisa di buka aplikasinya parah bnget gopay
Case Folding	tidak bisa di buka aplikasinya parah bnget gopay
Tokenizing	['Tidak', 'bisa', 'di', 'buka', 'aplikasinya', 'parah', 'bnget', 'gopay']
Normalization	['Tidak', 'bisa', 'di', 'buka', 'aplikasinya', 'parah', 'banget', 'gopay']
Stopword Removal	['buka', 'aplikasinya', 'parah', 'gopay']
Stemming	['buka', 'aplikasinya', 'parah', 'gopay']

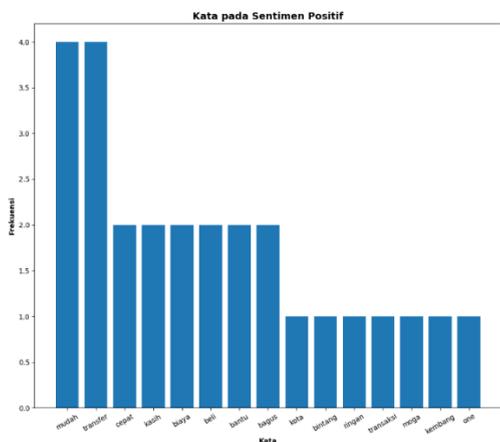
3. Transformation

Tahap transformation ini diawali dengan melakukan pembagian data menjadi data training dan data testing. Pembagian data ini menggunakan fungsi train test split dengan parameter random_state=42. Tabel 2 menunjukkan hasil pembagian data.

Table 2. Hasil Pembagian Data

Skenario Rasio Perbandingan	Data Training	Data Testing
80 : 20	23643	5910

Setelah data terbagi menjadi data training dan data testing, selanjutnya dilakukan pembobotan kata dengan metode TF-IDF. Gambar 3 merupakan contoh hasil pembobotan kata dengan TF-IDF.



Gambar 3 Hasil Kata Sentimen

4. Data Mining

Pada tahap ini, proses data mining melibatkan klasifikasi sentimen pada data yang telah dibagi ke dalam 10 fold menggunakan kernel linear pada algoritma Support Vector Machine (SVM). SVM merupakan salah satu metode machine learning yang umum digunakan untuk klasifikasi dan regresi.

Hasil dari skenario tersebut menunjukkan tingkat akurasi klasifikasi sentimen pada masing-masing fold. Akurasi diukur sebagai persentase keberhasilan model dalam memprediksi sentimen dengan benar. Berikut adalah rincian hasil akurasi untuk masing-masing fold:

Table 3. Hasil Klasifikasi dengan SVM

Fold	Kernel	Accuracy
Fold 1	Linear	100%
Fold 2		100%
Fold 3		100%
Fold 4		50%
Fold 5		0%
Fold 6		100%
Fold 7		0%
Fold 8		50%
Fold 9		50%
Fold 10		100%
Mean		65%

Dari hasil tersebut, terlihat variasi akurasi antar fold, yang mungkin disebabkan oleh perbedaan karakteristik data pada setiap fold. Evaluasi kinerja model pada setiap fold sangat penting untuk memahami sejauh mana model dapat digeneralisasi pada data yang berbeda.

Penggunaan kernel linear pada SVM menunjukkan performa yang cukup baik dalam tugas klasifikasi sentimen pada dataset ini. Dengan menganalisis hasil ini, dapat dilakukan pengoptimalan lebih lanjut pada model atau eksplorasi metode lain untuk meningkatkan performa klasifikasi sentimen pada data yang digunakan.

5. Evaluation

Table 4. Hasil Evaluasi

Accuracy	Precision	Recall	F1 Score
65%	71,43%	50%	58,82%

Dalam tahap evaluasi, kinerja model klasifikasi sentimen diukur menggunakan beberapa metrik evaluasi klasifikasi, yaitu Akurasi (Accuracy), Presisi (Precision), Recall, dan F1 Score. Berikut adalah rincian hasil evaluasi untuk model tersebut:

- Akurasi (Accuracy): 65%
Akurasi mengukur sejauh mana model dapat memprediksi sentimen dengan benar secara keseluruhan. Dalam konteks ini, model mencapai akurasi sebesar 65%
- Presisi (Precision): 71,43%
Presisi mengukur sejauh mana prediksi positif model benar. Dalam hal ini, 71,43% dari sentimen yang diprediksi positif oleh model memang benar-benar positif.

- InFormASi Dan Ilmu KOMputer), 14(1), 75–79. <https://doi.org/https://doi.org/10.37859/jf.v14i1.6795>
- Benner, M.J., Tushman, M.L., 2003. Exploitation, exploration, and process management: the productivity dilemma revisited. *Academy of Management Review* 28 (2), 238–256
- Fatkhudin, A., Artanto, F. A., Safli, N. A., & Wibowo, D. (2024). Decision Tree Berbasis SMOTE dalam Analisis Sentimen Penggunaan Artificial Intelligence untuk Skripsi. *Remik: Riset Dan E-Jurnal Manajemen Informatika Komputer*, 8(April), 494–505.
- Fatkhudin, A., Febrianto, M. Y., Artanto, F. A., Hadinata, M. W. N., & Fahlevi, R. (2022). Algoritma Decision Tree C.45 dalam analisa kelulusan mahasiswa Program Studi Manajemen Informatika UMPP. *Jurnal Ilmiah Ilmu Komputer Fakultas Ilmu Komputer Universitas Al Asyariah Mandar*, 8(2), 83–86.
- Fatkhudin, A., Khambali, A., Artanto, F. A., & Zade, N. A. P. (2023). Implementasi Algoritma Clustering K-Means Dalam Pengelompokan Mahasiswa Studi Kasus (Prodi Manajemen Informatika). *Jurnal Minfo Polgan*, 12(2), 777–783. <https://doi.org/10.33395/jmp.v12i2.12494>
- Nurian, A. (2023). Analisis Sentimen Ulasan Pengguna Aplikasi Google Play Menggunakan Naïve Bayes. *Jurnal Informatika dan Teknik Elektro Terapan*, 11(3s1).
- Rosyadi, I., Artanto, F. A., Rahmawati, S. E., Tri, H., & Joyo, B. (2022). Decision Tree Dalam Analisis Keputusan Pembelian Program Pada Perkumpulan Penggiat Programmer Indonesia. *Jurnal Fasilkom*, XII(III), 141–144.
- Rosyadi, I., Kusumawardani, H. H., Artanto, F. A., Hardani, A. A. A., & Nafilaturrosyidah, F. (2023). Clustering K-Means Dalam Pengelompokan Penjualan Produk Pada RTO Group. *Teknomatika*, 13(02), 55–60. <http://ojs.palcomtech.ac.id/index.php/teknomatika/article/view/618/439>
- Sondakh, D. E., Kom, S., Taju, S. W., Tene, M. G., & Pangaila, A. E. (2023). Sistem Analisis Sentimen Ulasan Aplikasi Belanja Online Menggunakan Metode Ensemble Learning. *CogITo Smart Journal*, 9(2), 280-2