

Prediksi Employee Churn Dengan Uplift Modeling Menggunakan Algoritma Logistic Regression

*Jovan Kinoto¹, Jansen Liharma Damanik², Erwin Tri Saputra Situmorang³, Josua Siregar⁴, Mawaddah Harahap⁵

Address: Universitas Prima Indonesia, Fakultas Teknologi dan Ilmu Komputer, Program Studi Teknik Informatika^{1,2,3,4,5}

Email: kinotojovan@gmail.com¹, janseliharma7@gmail.com², erwinstmrg16@gmail.com³, siregarjosua22@gmail.com⁴, mawaddah@unprimdn.ac.id⁵

Abstrak

Pada sebuah perusahaan, karyawan merupakan aset yang berharga dan dapat menunjang kesuksesan perusahaan tersebut. Namun, hilangnya tenaga kerja dapat merugikan perusahaan. Kondisi ini disebut dengan Employee Churn. Salah satu solusi untuk mengatasi Employee Churn adalah dengan menerapkan model Uplift Modeling. Dalam penelitian ini, penulis menganalisa penerapan Logistic Regression terhadap Uplift Modeling dalam permasalahan Employee Churn. Data yang diteliti adalah data karyawan dari IBM HR Analytics. Hasil prediksi pada penelitian ini mendapat akurasi sebesar 64,40%, sedangkan hasil preskripsi menghasilkan hasil yang cukup baik apabila menerapkan waktu kerja tambahan pada karyawan. Berdasarkan hasil yang didapat, diketahui bahwa para karyawan justru cenderung bertahan di perusahaan apabila diberikan waktu kerja tambahan.

Keywords: *Employee Churn, Uplift Modeling, Logistic Regression, Lai's Generalized Weighted Uplift Method*

1. Latar Belakang

Dalam sebuah perusahaan, karyawan merupakan salah satu bagian utama yang menjalankan perusahaan. Tanpa karyawan, perusahaan tidak akan dapat beroperasi sama sekali. Karyawan yang keluar akan diganti dengan karyawan baru, atau dapat disebut juga dengan pergantian karyawan atau *Employee Churn*. *Employee Churn* seringkali merugikan perusahaan, dan telah banyak peneliti yang meneliti permasalahan ini [1]. Kerugian yang disebabkan oleh pergantian karyawan antara lain biaya rekrutmen baru, pelatihan karyawan baru, dan potensi hilangnya karyawan berpengaruh [2]. Ada beberapa hal yang menyebabkan seorang karyawan untuk keluar dari perusahaan, seperti lingkungan kerja, lokasi kerja, kesetaraan *gender*, kesetaraan gaji, dan lain-lain [3]. Strategi yang diterapkan berbagai organisasi untuk mengurangi *Employee Churn* antara lain kenaikan gaji, kesempatan untuk jalan-jalan, rotasi pekerjaan yang sering, dan lain-lain [4].

Peneliti-peneliti sebelumnya telah meneliti data karyawan dengan berbagai model. Model-model tersebut antara lain *Resume Prediction* [5], [6], *Churn Prediction* [7], [8], dan *Uplift Modeling* [9], [10]. Ketiga model

tersebut menggunakan algoritma *Machine Learning* yang berbeda-beda, seperti *Random Forest*, *Neural Network*, dan *Extreme Gradient Boosting*. Selain dari ketiga algoritma tersebut, *Logistic Regression* adalah salah satu algoritma *Machine Learning* yang juga dapat digunakan untuk memprediksi. Algoritma ini menggunakan teknik regresi. *Logistic Regression* juga telah digunakan pada penelitian sebelumnya untuk meneliti prakiraan hujan es [11], *customer churn* [12], [13], dan performa akademik [14].

Dalam penelitian ini, penulis menerapkan model *Uplift Modeling* yang membandingkan variabel *Treatment* dalam memprediksi. Model ini juga telah digunakan untuk meneliti pencegahan *dropout* [15], pemasaran [16], dan periklanan [17]. Penelitian ini menggunakan dataset *IBM HR Analytics* [18] yang juga telah diteliti oleh peneliti-peneliti sebelumnya [19]–[22]. Dataset ini berisi data pribadi dan pekerjaan karyawan. Penelitian ini menggunakan algoritma *Logistic Regression* dan 2 model *Uplift Modeling*. Diharapkan penelitian ini dapat menambah wawasan tentang model *Uplift* dalam memprediksi dan mempreskripsi *Employee Churn*.

A. Rumusan Masalah

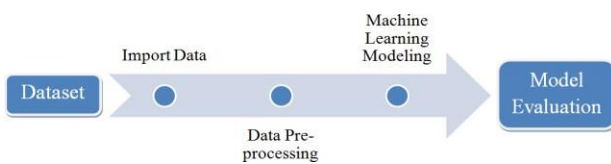
Berdasarkan latar belakang di atas, rumusan masalah dalam penelitian ini adalah bagaimana cara mengatasi *Employee Churn* menggunakan *Uplift Modeling* dan algoritma *Logistic Regression*.

B. Tujuan

Tujuan dari penelitian ini adalah untuk menganalisa penerapan model *Uplift Modeling* menggunakan algoritma *Logistic Regression* dalam permasalahan *Employee Churn*.

2. Metode

Dalam penelitian ini terdapat empat prosedur utama, antara lain proses *Import Data*, *Data Pre-processing*, *Machine learning Modeling*, dan *Model Evaluation*.



Gambar 2. Diagram Prosedur Kerja

a. *Import Data*

Dalam tahap ini, *dataset* diimpor ke dalam format *dataframe* agar dapat diproses pada tahap berikutnya. Paket *Pandas* digunakan untuk melakukan proses impor.

b. *Data Pre-processing*

Dalam tahap ini, *dataset* yang telah diimpor diolah terlebih dahulu agar dapat diproses algoritma *Machine Learning*. Tahap ini dibagi lagi menjadi beberapa bagian, antara lain sebagai berikut:

1. Pembersihan Data

Dataset umumnya mempunyai data-data yang tidak bernilai atau format yang tidak sesuai. Maka dari itu, pada tahap ini dilakukan “pembersihan” data agar *dataset* dapat diproses ke tahap selanjutnya. Terdapat dua cara yang digunakan pada penelitian ini, seperti *Label Encoding* dan *One-Hot Encoding* yang mengubah data non-numerik menjadi numerik.

2. *Feature Engineering*

Feature Engineering berfungsi untuk melakukan analitik preskriptif pada penelitian ini. Tahap ini dilakukan dengan menambahkan *feature* baru berupa *Target Class* agar kurva *Qini* yang dihasilkan dapat menampilkan nilai *Uplift*. Penampilan kurva *Qini* dilakukan dengan memperhatikan korelasi *Treatment*, karena terdapat dua macam *Treatment* yang dapat digunakan pada penelitian ini, yaitu antara lain yang berkorelasi positif dan negatif terhadap *Employee Churn*.

3. *Korelasi Treatment*

Penelitian ini menggunakan fungsi *Pearson Correlation* dari paket *Pandas* untuk menganalisis korelasi antara variabel *Treatment* dan *Employee Churn*. Rumus *Pearson Correlation* dapat ditulis sebagai berikut:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}} \quad (1)$$

Analisis ini harus dilakukan karena *dataset* dengan *Treatment* negatif mempunyai hasil yang berbeda dari *Treatment* positif saat diprediksi.

c. *Machine Learning Modeling*

Pada tahap ini dilakukan proses *Machine Learning Modeling*. Tahap-tahapnya adalah sebagai berikut:

1. *Data split*: tahap ini membagi *dataset* menjadi dua bagian, yaitu *Train Data* dan *Test Data*. Proporsi *Test Data* mengikuti penelitian sebelumnya [26], yakni sebanyak 1/3 dari jumlah keseluruhan data. Agar dipastikan proporsi *Treatment* pada *Train Data* dan *Test Data* sama besar, digunakan *Stratify* pada kolom *Treatment*.
2. *Model training*: tahap ini melatih model dengan algoritma *Logistic Regression* menggunakan *Train Data* yang telah dibagi sebelumnya.
3. *Data testing*: tahap ini menguji model yang telah dilatih sebelumnya dengan *Test Data* dan mendapatkan hasil prediksi dan probabilitas.

d. *Model Evaluation*

Pada penelitian ini, metode *Model Evaluation* yang digunakan terbagi menjadi dua antara lain, metrik akurasi untuk mengevaluasi hasil prediksi dan Kurva *Qini* untuk mengevaluasi nilai *Uplift* yang berhasil diprediksi. *Confusion Matrix* juga ditampilkan pada hasil prediksi.

Akurasi model *Uplift* dan *Confusion Matrix* dihitung menggunakan *Scikit-Learn*. *Confusion Matrix* mempunyai empat parameter, antara lain *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)*.

Perhitungan akurasi model *Uplift* menggunakan empat keluaran yang terdiri dari *Control Non-responder (CN)*, *Control Responder (CR)*, *Treatment Non-responder (TN)*, dan *Treatment Responder (TR)* yang disimbolkan dengan nilai 0, 1, 2, dan 3 secara berurutan.. Kemudian, jika *i* adalah keluaran yang akan diprediksi dan TP, TN, FP, dan FN adalah parameter *Confusion Matrix*, maka rumus akurasi model *Uplift* dapat ditulis sebagai berikut:

$$\text{Akurasi model Uplift} = \frac{\sum_{i=0}^3 \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}}{4} \times 100\% \quad (2)$$

Kurva *Qini* pada penelitian ini ditampilkan menggunakan paket *Matplotlib* dan *Seaborn*. Jika diketahui total populasi data karyawan dalam sampel (*N*), proporsi yang ditarget dengan *Employee Retention Strategy* (α), dan jumlah populasi dengan *Target Class* tertentu (*n*), maka rumus kurva *Qini* untuk menghitung nilai *Uplift* dapat ditulis sebagai berikut:

$$\text{Kurva Qini } (\alpha) = N\alpha \left[\frac{n(\text{TR})}{n(\text{C})} - \frac{n(\text{TR}) \cdot n(\text{T})}{n(\text{C}) \cdot n(\text{T})} \right] \quad (3)$$

Rumus di atas kemudian dinormalisasi ke dalam bentuk persentase menjadi:

$$\text{Kurva Qini } (\alpha) = N\alpha \left[\frac{n(\text{TR})}{n(\text{T})} - \frac{n(\text{CR})}{n(\text{C})} \right] \quad (4)$$

Jika kurva *Uplift* yang dihasilkan secara keseluruhan lebih tinggi dari kurva *Random Model*, maka analitik preskriptif model tersebut dinyatakan berhasil. *Random Model* adalah garis penghubung nilai awal dan akhir pada kurva *Qini* dan menandakan bahwa nilai *Uplift* yang dihasilkan tidak menggunakan model apapun.

3. Bahan

Dataset yang digunakan pada penelitian ini adalah *dataset* dari IBM dengan nama *IBM HR Analytics Employee Attrition & Performance* [18]. *Dataset* ini memiliki 1.470 data karyawan dan 35 variabel. Penelitian sebelumnya sudah dilakukan dengan *dataset* ini menggunakan model CP [19]–[21]. Variabel *Treatment* pada *dataset* ini adalah *overtime* (jam kerja yang melebihi seharusnya).

4. Hasil

Pengujian model dilakukan pada software Jupyter Notebook. Berikut ini hasil penelitian pada setiap tahap penelitian.

a. Hasil Import Data

Tahap pertama yang dilakukan adalah mengimpor *dataset* ke dalam Jupyter Notebook dan mengkonversi format dari CSV (Comma-Separated Values) ke *dataframe* dengan menggunakan paket *Pandas*. *Dataset* ini berisi tentang data pribadi dan pekerjaan karyawan. Variabel yang dipilih untuk *Employee Retention Strategy (ERS)* adalah *Overtime*.

b. Hasil Pembersihan Data

Dataset belum dapat diproses oleh *machine learning* dikarenakan masih terdapat data yang bukan berupa angka. Data tersebut diproses terlebih dahulu menggunakan *Label Encoding* dan *One-Hot Encoding*. *Label Encoding* mengubah variabel *Treatment* menjadi binary (1 dan 0). *One-hot Encoding* mengubah variabel yang mempunyai beberapa parameter menjadi variabel baru dengan tipe data binary. Sedangkan untuk variabel yang tidak bernilai (null) tidak ditemukan, tetapi variabel dengan satu parameter dihapus karena tidak dapat digunakan. Hasil pembersihan data dijelaskan lebih mendetail pada tabel di bawah ini.

Tabel 1 Penerapan Pre-processing pada data yang bukan angka

No	Nama Variabel	Masalah	Solusi
1	Department	Banyak parameter	One-Hot Encoding
2	EducationField	Tipe data object	One-Hot Encoding
3	EmployeeCount	Parameter sama	Drop

4	EmployeeNumber	Parameter sama	Drop
5	Gender	Banyak parameter	One-Hot Encoding
6	JobRole	Banyak parameter	One-Hot Encoding
7	MaritalStatus	Banyak parameter	One-Hot Encoding
8	Over18	Parameter sama	Drop
9	OverTime	Variabel Treatment	Label Encoding ("Yes" = 0, "No" = 1)
10	StandardHours	Parameter sama	Drop

c. Hasil Feature Engineering

Feature Engineering menghasilkan variabel *Target Class* sesuai dengan tabel di bawah ini.

Tabel 2 Ketentuan dalam penilaian variabel *Target Class*

		Nilai Employee Churn	
		1	0
Nilai Treatment	0	0 (CN)	1 (CR)
	1	2 (TN)	3 (TR)

Variabel *Target Class* digunakan untuk membuat kurva *Qini* dan sebagai variabel dependen untuk model *Uplift*.

d. Hasil Korelasi Treatment

Korelasi *Treatment* pada *dataset* dihitung dengan menggunakan fungsi *Pearson Correlation* dari paket *Pandas* dan menghasilkan -24,61%. Apabila nilai variabel tersebut dibalik (nilai 0 menjadi 1, dan nilai 1 menjadi 0), maka akan menghasilkan variabel *Treatment* yang bernilai positif (*Not Overtime*).

e. Hasil Machine Learning Modeling

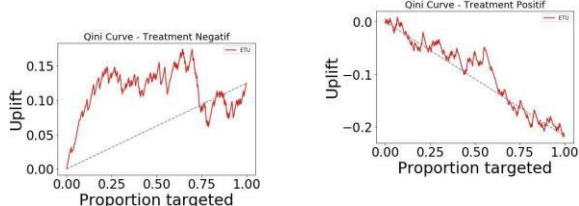
Tahap ini terbagi menjadi hasil *data split* dan hasil prediksi.

1. Hasil Data Split

Dataset dibagi menjadi *Train Data* dan *Test Data* sebanyak 1.029 dan 441 data secara berurutan menggunakan *Scikit-Learn*. *Dataset* dibagi dengan menggunakan parameter *Stratify* pada variabel *Treatment* agar pembagian data lebih seimbang.

2. Hasil Prediksi

Model *Uplift* dilatih terlebih dahulu dengan *Train Data* yang sudah dibagi sebelumnya. Setelah itu, model *Uplift* kemudian digunakan untuk memprediksi *Test Data*.



f. Hasil Evaluasi Prediksi

Tabel 3 Hasil Evaluasi Prediksi

			True	False
Treatment Negatif	CN	Positive	412	0
		Negative	29	0
	CR	Positive	345	0
		Negative	96	0
	TN	Positive	407	0
		Negative	34	0
	TR	Positive	0	159
		Negative	0	282
	Akurasi		63,95%	
	Treatment Positif	CN	Positive	411
Negative			30	0
CR		Positive	0	155
		Negative	0	286
TN		Positive	402	0
		Negative	39	0
TR		Positive	355	0
		Negative	86	0
Akurasi		64,85%		

Tabel di atas adalah hasil prediksi model *Uplift* menggunakan *Confusion Matrix* dan *Accuracy Score* dari paket *Scikit-Learn*. *True Positive* dan *True Negative* merupakan prediksi yang benar, sedangkan *False Positive* dan *False Negative* merupakan prediksi yang salah. *Treatment* positif diperoleh dengan membalikkan nilai pada variabel *Treatment* dari 0 ke 1 dan sebaliknya. Hasil

akurasi yang didapatkan dari *Treatment* negatif dan positif secara berurutan adalah 63,95% dan 64,85%.

g. Hasil Evaluasi Preskripsi

Evaluasi preskripsi dihasilkan menggunakan kurva *Qini* dari paket *Matplotlib* dan *Seaborn*. Kurva *Qini* yang dihasilkan adalah sebagai berikut:

Gambar 2. Kurva *Qini* Hasil Preskripsi

Garis berwarna merah menyatakan model *Uplift* dengan algoritma *Logistic Regression*, sedangkan garis titik diagonal berwarna abu-abu menyatakan model *Random* (tidak menggunakan model apapun). Berdasarkan kurva-kurva di atas, model *Uplift* berhasil mempreskripsi kedua *Treatment*. Hal ini dapat dilihat dari garisberwarna merah yang secara keseluruhan berada di atas garis abu-abu.

5. Pembahasan

Kedua *Treatment* dalam penelitian ini mempunyai hasil yang sangat mirip. *Treatment* positif (*Not Overtime*) mempunyai akurasi 0,90% lebih tinggi dari *Treatment* negatif (*Overtime*). Jika dibandingkan dengan model *Uplift* yang menggunakan algoritma lain, seperti *Extreme Gradient Boosting* (XGB) [10], model ini masih memiliki akurasi yang lebih tinggi. Namun dalam hasil preskripsi, algoritma XGB secara keseluruhan memiliki nilai *Uplift* yang lebih baik dari algoritma *Logistic Regression*. Meskipun memiliki tingkat akurasi yang tidak terlalu akurat, kedua *Treatment* yang diteliti pada penelitian ini berhasil memberikan hasil yang cukup baik dalam mengurangi *Employee Churn*. Namun, apabila dibandingkan yang mana yang lebih baik, pemberian waktu kerja lebih pada karyawan (*Overtime*) mempunyai hasil yang lebih baik dalam mengurangi *Employee Churn* dibandingkan dengan tidak memberikan waktu kerja lebih pada karyawan (*Not Overtime*). Hal ini mengindikasikan para karyawan yang diteliti lebih senang apabila bekerja lembur.

6. Kesimpulan

Penelitian ini berhasil memprediksi dan mempreskripsi permasalahan *Employee Churn* dengan *Uplift Modeling* menggunakan algoritma *Logistic Regression* pada dataset IBM HR Analytics. Variabel *Treatment* yang digunakan berupa variabel *Overtime* dan *Not Overtime*. Meskipun model yang diteliti hanya memiliki total akurasi sebesar 64,40%. Namun, model tersebut menghasilkan hasil yang cukup baik dalam mengurangi *Employee Churn*, dan apabila dibandingkan maka variabel *Overtime* lebih unggul dibandingkan dengan *Not Overtime*. Sehingga, dapat disimpulkan bahwa para karyawan yang diteliti lebih senang apabila bekerja lembur.

Ada beberapa saran yang dapat diberikan penulis untuk penelitian selanjutnya, antara lain sebagai berikut:

1. Penelitian dengan algoritma atau dataset yang berbeda,
2. Membandingkan algoritma Logistic Regression dengan algoritma lainnya untuk mencari algoritma yang paling efektif.

UCAPAN TERIMA KASIH

1. Ibu Dr. Chrismis Novalinda Ginting, S.SiT., M.Kes. selaku Rektor Universitas Prima Indonesia.
2. Bapak Mardi Turnip, S.Kom., M.Kom. selaku Dekan Fakultas Teknologi dan Ilmu Komputer Universitas Prima Indonesia.
3. Ibu Mawaddah Harahap, S.Kom., M.Kom. selaku Ketua Program Studi S-1 Teknik Informatika Fakultas Teknologi dan Ilmu Komputer Universitas Prima Indonesia dan Dosen Pembimbing yang telah meluangkan waktu untuk memberikan bimbingan dalam penyusunan laporan penelitian ini.
4. Bapak Allwin M. Simarmata, S.Kom., M.Kom. selaku Dosen Pengulas yang telah mengulas dan memberikan arahan dalam penyusunan laporan penelitian ini.
5. Seluruh Dosen dan Staff Program Studi S-1 Teknik Informatika Fakultas Teknologi dan Ilmu Komputer Universitas Prima Indonesia.
6. Orang tua yang telah memberikan berbagai dukungan dan semangat.

References

- [1] S. Kim, L. Tam, J.-N. Kim, and Y. Rhee, "Determinants of employee turnover intention," *Corp. Commun. An Int. J.*, 2017.
- [2] D. S. Sisodia, S. Vishwakarma, and A. Pujahari, "Evaluation of machine learning models for employee churn prediction," in 2017 *International Conference on Inventive Computing and Informatics (ICICI)*, 2017, pp. 1016–1020.
- [3] İ. O. Yiğit and H. Shourabizadeh, "An approach for predicting employee churn by using data mining," in 2017 *International Artificial Intelligence and Data Processing Symposium (IDAP)*, 2017, pp. 1–4.
- [4] D. K. Srivastava and P. Nair, "Employee attrition analysis using predictive techniques," in *International Conference on Information and Communication Technology for Intelligent Systems*, 2017, pp. 293–300.
- [5] J. M. Kirimi and C. A. Moturi, "Application of data mining classification in employee performance prediction," *Int. J. Comput. Appl.*, vol. 146, no. 7, pp. 28–35, 2016.
- [6] M. He, D. Shen, Y. Zhu, R. He, T. Wang, and Z. Zhang, "Career Trajectory Prediction based on CNN," in 2019 *IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI)*, 2019, pp. 22–26.
- [7] S. H. Dolatabadi and F. Keynia, "Designing of customer and employee churn prediction model based on data mining method and neural predictor," in 2017 *2nd International Conference on Computer and Communication Systems (ICCCS)*, 2017, pp. 74–77.
- [8] A. Alamsyah and N. Salma, "A Comparative Study of Employee Churn Prediction Model," in 2018 *4th International Conference on Science and Technology (ICST)*, 2018, pp. 1–4.
- [9] E. Rombaut and M.-A. Guerry, "The effectiveness of employee retention through an uplift modeling approach," *Int. J. Manpow.*, 2020.
- [10] D. Wijaya, J. H. Ds, B. Pasaribu, S. Barus, and A. Dharma, "Uplift modeling VS conventional predictive model : A reliable machine learning model to solve employee turnover," *Int. J. Artif. Intell. Res.*, vol. 4, no. 2, 2020.
- [11] J. L. Sánchez, J. L. Marcos, M. T. De la Fuente, and A. Castro, "A logistic regression model applied to short term forecast of hail risk," *Phys. Chem. Earth*, vol. 23, no. 5–6, pp. 645–648, 1998.
- [12] A. De Caigny, K. Coussement, and K. W. De Bock, "A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees," *Eur. J. Oper. Res.*, vol. 269, no. 2, pp. 760–772, 2018.
- [13] P. K. Dalvi, S. K. Khandge, A. Deomore, A. Bankar, and V. A. Kanade, "Analysis of customer churn prediction in telecom industry using decision trees and logistic regression," in 2016 *Symposium on Colossal Data Analysis and Networking (CDAN)*, 2016, pp. 1–4.
- [14] M. E. Urrutia-Aguilar, R. Fuentes-García, V. D. M. Martínez, E. Beck, S. O. León, and R. Guevara-Guzmán, "Logistic regression model for the academic performance of first-year medical students in the biomedical area," *Creat. Educ.*, vol. 7, no. 15, p. 2202, 2016.
- [15] D. Olaya, J. Vásquez, S. Maldonado, J. Miranda, and W. Verbeke, "Uplift Modeling for preventing student dropout in higher education," *Decis. Support Syst.*, p. 113320, 2020.
- [16] P. Rzepakowski and S. Jaroszewicz, "Uplift modeling in direct marketing," *J. Telecommun. Inf. Technol.*, pp. 43–50, 2012.
- [17] S. Kawanaka and D. Moriwaki, "Uplift modeling for location-based online advertising," in *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-based Recommendations, Geosocial Networks and Geoadvertising*, 2019, pp. 1–4.

- [18] Pavansubhash, "IBM HR Analytics Employee Attrition & Performance," Kaggle - Database: Open Database, Contents: Database Contents, 2017. <https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-Dataset> (accessed Oct. 09, 2020).
- [19] M. Maisuradze, "Predictive Analysis On The Example Of Employee Turnover," Tallinn Univ. Technol., 2017.
- [20] K. Sehgal, H. Bindra, A. Batra, and R. Jain, "Prediction of Employee Attrition Using GWO and PSO Optimised Models of C5. 0 Used with Association Rules and Analysis of Optimisers," in *Innovations in Computer Science and Engineering*, Springer, 2019, pp. 1–8.
- [21] S. S. Alduayj and K. Rajpoot, "Predicting employee attrition using machine learning," in *2018 International Conference on Innovations in Information Technology (IIT)*, 2018, pp. 93–98.
- [22] K. G. King, "Data analytics in human resources: A case study and critical review," *Hum. Resour. Dev. Rev.*, vol. 15, no. 4, pp. 487–495, 2016.
- [23] A. Shaar, T. Abdessalem, and O. Segard, "Pessimistic uplift modeling," 2016.
- [24] A. Betlei, E. Diemert, and M.-R. Amini, "Uplift prediction with dependent feature representation in imbalanced treatment and control conditions," in *International Conference on Neural Information Processing*, 2018, pp. 47–57.
- [25] R. Gubela, A. Bequé, S. Lessmann, and F. Gebert, "Conversion uplift in ecommerce: A systematic benchmark of modeling strategies," *Int. J. Inf. Technol. Decis. Mak.*, vol. 18, no. 03, pp. 747–791, 2019.
- [26] F. Devriendt, J. Berrevoets, and W. Verbeke, "Why you should stop predicting customer churn and start using uplift models," *Inf. Sci. (Ny)*, 2019.