

**Ballings, Michel; Van den Poel, Dirk**

**CRM in social media: predicting increases in Facebook usage frequency.** (English)

Zbl 1346.90412

Eur. J. Oper. Res. 244, No. 1, 248-260 (2015).

**Summary:** The purpose of this study is to (1) assess the feasibility of predicting increases in Facebook usage frequency, (2) evaluate which algorithms perform best, (3) and determine which predictors are most important. We benchmark the performance of Logistic Regression, Random Forest, Stochastic Adaptive Boosting, Kernel Factory, Neural Networks and Support Vector Machines using five times twofold cross-validation. The results indicate that it is feasible to create models with high predictive performance. The top performing algorithm was Stochastic Adaptive Boosting with a cross-validated AUC of 0.66 and accuracy of 0.74. The most important predictors include deviation from regular usage patterns, frequencies of likes of specific categories and group memberships, average photo album privacy settings, and recency of comments. Facebook and other social networks alike could use predictions of increases in usage frequency to customize its services such as pacing the rate of advertisements and friend recommendations, or adapting News Feed content altogether. The main contribution of this study is that it is the first to assess the prediction of increases in usage frequency in a social network.

**MSC:**

90B50 Management decision making, including multiple objectives

Cited in 8 Documents

**Keywords:**

decision support systems; social media; data mining; predictive analytics; Facebook

**Software:**

MASS (R); R; LIBSVM; randomForest; glmnet

**Full Text:** DOI

**References:**

- [1] Alpaydin, E., Combined  $5 \times 2$  cv  $\textit{testit}\{f\}$  test for comparing supervised classification learning algorithms, *Neural Computation*, 11, 8, 1885-1892, (1999)
- [2] Aral, S.; Muchnik, L.; Sundararajan, A., Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks, *Proceedings of the National Academy of Sciences*, 106, 21544-21549, (2009)
- [3] Aral, S.; Walker, D., Creating social contagion through viral product design: A randomized trial of peer influence in networks, *Management Science*, 57, 9, 1623-1639, (2011)
- [4] Arnaboldi, V.; Guazzini, A.; Passarella, A., Egocentric online social networks: analysis of key features and prediction of tie strength in facebook, *Computer Communications*, 36, 10-11, 1130-1144, (2013)
- [5] Au, W. H.; Chan, K. C.C.; Yao, X., A novel evolutionary data mining algorithm with applications to churn prediction, *IEEE Trans. Evol. Comput.*, 7, 6, 532-545, (2003)
- [6] Baecke, P.; Van den Poel, D., Improving customer acquisition models by incorporating spatial autocorrelation at different levels of granularity, *Journal of Intelligent Information Systems*, 41, 1, 73-90, (2013)
- [7] Baesens, B.; Verstraeten, G.; Van den Poel, D.; Egmont-Petersen, M.; Van Kenhove, P.; Vanthienen, J., Bayesian network classifiers for identifying the slope of the customer lifecycle of long-life customers, *European Journal of Operational Research*, 156, 2, 508-523, (2004) · Zbl 1056.90019
- [8] Bakshy, E.; Eckles, D.; Yan, R.; Rosenn, I., Social influence in social advertising: evidence from field experiments, *Proceedings of the 13th ACM conference on electronic commerce*, 146-161, (2012), New York
- [9] Ballings, M.; Van den Poel, D., Customer event history for churn prediction: how long is long enough?, *Expert Systems with Applications*, 39, 18, 13517-13522, (2012)
- [10] Ballings, M.; Van den Poel, D., Kernel factory: an ensemble of kernel machines, *Expert Systems with Applications*, 40, 8, 2904-2913, (2013)
- [11] Ballings, M. Van den Poel, D. (2013b). Kernel Factory: an ensemble of kernel machines. R package version 0.1.1.
- [12] Bauer, E.; Kohavi, R., An empirical comparison of voting classification algorithms: bagging, boosting, and variants, *Machine*

- Learning, 36, 1-2, 105-139, (1999)
- [13] Ben-Hur, A.; Weston, J., A user's guide to support vector machines, *Methods in molecular biology*, 223-239, (2010), Department of Computer Science Colorado State University
- [14] Benoit, D. F.; Van den Poel, D., Improving customer retention in financial services using kinship network information, *Expert Systems with Applications*, 39, 13, 11435-11442, (2012)
- [15] Berkson, J., Application of the logistic function to bio-assay, *Journal of the American Statistical Association*, 39, 227, 357-365, (1944)
- [16] Bishop, C., *Neural Networks for Pattern Recognition*, (2002), Oxford University Press
- [17] Bolton, R. N., A dynamic model of the duration of the customer's relationship with a continuous service provider: the role of satisfaction, *Marketing Science*, 17, 1, 45, (1998)
- [18] Breiman, L., Random forests, *Machine Learning*, 45, 1, 5-32, (2001) · [Zbl 1007.68152](#)
- [19] Buckinx, W.; Van den Poel, D., Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting, *European Journal of Operational Research*, 164, 1, 252-268, (2005) · [Zbl 1132.90349](#)
- [20] Burez, J.; Van den Poel, D., CRM at a pay-TV company: using analytical models to reduce customer attrition by targeted marketing for subscription services, *Expert Systems with Applications*, 32, 2, 277-288, (2007)
- [21] Burez, J.; Van den Poel, D., Separating financial from commercial customer churn: A modeling step towards resolving the conflict between the sales and credit department, *Expert Systems with Applications*, 35, 1-2, 497-514, (2008)
- [22] Burez, J.; Van den Poel, D., Handling class imbalance in customer churn prediction, *Expert Systems with Applications*, 36, 3, 4626-4636, (2009)
- [23] Claussen, J.; Kretschmer, T.; Mayrhofer, P., The effects of rewarding user engagement: the case of facebook apps, *Information Systems Research*, 24, 1, 186-200, (2013)
- [24] Clayton, R. B.; Osborne, R. E.; Miller, B. K.; Oberle, C. D., Loneliness, anxiousness, and substance use as predictors of facebook use, *Computers in Human Behavior*, 29, 3, 687-693, (2013)
- [25] Cortes, C.; Vapnik, V., Support-vector networks, *Machine Learning*, 20, 3, 273-297, (1995) · [Zbl 0831.68098](#)
- [26] Coussemont, K.; Benoit, D. F.; Van den Poel, D., Improved marketing decision making in a customer churn prediction context using generalized additive models, *Expert Systems with Applications*, 37, 3, 2132-2143, (2010)
- [27] Coussemont, K.; Van den Poel, D., Churn prediction in subscription services: an application of support vector machines while comparing two parameter-selection techniques, *Expert Systems with Applications*, 34, 1, 313-327, (2008)
- [28] Coussemont, K.; Van den Poel, D., Improving customer attrition prediction by integrating emotions from client/company interaction emails and evaluating multiple classifiers, *Expert Systems with Applications*, 36, 3, 6127-6134, (2009)
- [29] Culp, M., Johnson, K. Michailidis, G. (2012). *ada: an R package for stochastic boosting*. R package version 2.0-3.
- [30] Datta, P.; Masand, B.; Mani, D. R.; Li, B., Automated cellular modeling and prediction on a large scale, *Artificial Intelligence Review*, 14, 6, 485-502, (2000) · [Zbl 0984.68546](#)
- [31] De Bock, K. W.; Van den Poel, D., An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction, *Expert Systems with Applications*, 38, 10, 12293-12301, (2011)
- [32] De Bock, K. W.; Van den Poel, D., Reconciling performance and interpretability in customer churn prediction using ensemble learning based on generalized additive models, *Expert Systems with Applications*, 39, 8, 6816-6826, (2012)
- [33] Demsar, J., Statistical comparisons of classifiers over multiple data sets, *Journal of Machine Learning Research*, 7, 1-30, (2006) · [Zbl 1222.68184](#)
- [34] Dietterich, T. G., Approximate statistical tests for comparing supervised classification learning algorithms, *Neural Computation*, 10, 7, 1895-1923, (1998)
- [35] Dreiseitl, S.; Ohno-Machado, L., Logistic regression and artificial neural network classification models: A methodology review, *Journal of Biomedical Informatics*, 35, 5-6, 352-359, (2002)
- [36] Eiben, A. E., Koudijs, A. E. Slisser, F. (1998). Genetic modelling of customer retention. In *Genetic programming. First European workshop, EuroGP'98 proceedings* (pp. 178-186).
- [37] Ellison, N. B.; Steinfield, C.; Lampe, C., The benefits of facebook "friends": social capital and college students' use of online social network sites, *Journal of Computer-Mediated Communication*, 12, 4, 1, (2007)
- [38] Facebook (2013a). News feed FYI: A window into news feed. <https://www.facebook.com/business/news/News-Feed-FYI-A-Window-Into-News-Feed/>.
- [39] Facebook (2013b). News feed FYI: More relevant ads in news feed | facebook newsroom. <http://newsroom.fb.com/news/2013/09/news-feed-fyi-more-relevant-ads-in-news-feed/>.
- [40] Facebook (2014). Newsroom - Key facts. <http://newsroom.fb.com/Key-Facts>.
- [41] Freund, Y.; Schapire, R., Experiments with a new boosting algorithm, *Machine learning. proceedings of the thirteenth international conference (ICML '96)*, 148-156, (1996), Bari, Italy
- [42] Friedman, J. H., Greedy function approximation: A gradient boosting machine, *Annals of Statistics*, 29, 5, 1189-1232, (2001) · [Zbl 1043.62034](#)
- [43] Friedman, J. H., Stochastic gradient boosting, *Computational Statistics & Data Analysis*, 38, 4, 367-378, (2002) · [Zbl 1072.65502](#)
- [44] Friedman, J.; Hastie, T.; Tibshirani, R., Regularization paths for generalized linear models via coordinate descent, *Journal of*

- Statistical Software, 33, 1, 1-22, (2010)
- [45] Friedman, J., Hastie, T. Tibshirani, R. (2013). glmnet: Lasso and elastic-regularized generalized linear models. R package version 1.9-3.
- [46] Gladly, N.; Baesens, B.; Croux, C., Modeling churn using customer lifetime value, *European Journal of Operational Research*, 197, 1, 402-411, (2009) · [Zbl 1157.91396](#)
- [47] Gómez, D.; Figueira, J. R.; Eusébio, A., Modeling centrality measures in social network analysis using bi-criteria network flow optimization problems, *European Journal of Operational Research*, 226, 2, 354-365, (2013) · [Zbl 1292.91149](#)
- [48] Guisan, A.; Edwards, T. C.; Hastie, T., Generalized linear and generalized additive models in studies of species distributions: setting the scene, *Ecological Modelling*, 157, 2-3, 89-100, (2002)
- [49] Hand, D. J., Good practice in retail credit scorecard assessment, *Journal of the Operational Research Society*, 56, 9, 1109-1117, (2005) · [Zbl 1097.91523](#)
- [50] Hanley, J.; Mcneil, B., The meaning and use of the area under a receiver operating characteristic (ROC) curve, *Radiology*, 143, 1, 29-36, (1982)
- [51] Hartmann, W. R.; Manchanda, P.; Nair, H.; Bothner, M.; Dodds, P.; Godes, D., Modeling social interactions: identification, empirical methods and policy implications, *Marketing Letters*, 19, 3-4, 287-304, (2008)
- [52] Hellmann, T.; Staudigl, M., Evolution of social networks, *European Journal of Operational Research*, 234, 3, 583-596, (2014) · [Zbl 1304.91168](#)
- [53] Hsu, C.-W.; Chang, C.-C.; Lin, C.-J., A practical guide to support vector classification, Technical Report, (2010), Department of Computer Science and Information Engineering, National Taiwan University Taipei 106, Taiwan
- [54] Hung, S.-Y.; Yen, D. C.; Wang, H.-Y., Applying data mining to telecom churn management, *Expert Systems with Applications*, 31, 3, 515-524, (2006)
- [55] Hwang, H.; Jung, T.; Suh, E., An LTV model and customer segmentation based on customer value: A case study on the wireless telecommunication industry, *Expert Systems with Applications*, 26, 2, 181-188, (2004)
- [56] Janitza, S.; Strobl, C.; Boulesteix, A.-L., An AUC-based permutation variable importance measure for random forests, *BMC Bioinformatics*, 14, 1, 119, (2013)
- [57] Kim, Y., Toward a successful CRM: variable selection, sampling, and ensemble, *Decision Support Systems*, 41, 2, 542-553, (2006)
- [58] King, R.; Feng, C.; Sutherland, A., Statlog - comparison of classification algorithms on large real-world problems, *Applied Artificial Intelligence*, 9, 3, 289-333, (1995)
- [59] Kumar, D. A.; Ravi, V., Predicting credit card customer churn in banks using data mining, *International Journal of Data Analysis Techniques and Strategies*, 1, 1, 4-28, (2008)
- [60] Lampe, C.; Ellison, N.; Steinfield, C., A familiar face(book): profile elements as signals in an online social network, *Proceedings of the conference on human factors in computing systems. Association for computing machinery*, 435-444, (2007), San Jose, CA
- [61] Larivière, B.; Van den Poel, D., Predicting customer retention and profitability by using random forests and regression forests techniques, *Expert Systems with Applications*, 29, 2, 472-484, (2005)
- [62] Lemmens, A.; Croux, C., Bagging and boosting classification trees to predict churn, *Journal of Marketing Research (JMR)*, 43, 2, 276-286, (2006)
- [63] Lemon, K. N.; White, T. B.; Winer, R. S., Dynamic customer relationship management: incorporating future considerations into the service retention decision, *Journal of Marketing*, 66, 1, 1-14, (2002)
- [64] Lewis, K.; Kaufman, J.; Gonzalez, M.; Wimmer, A.; Christakis, N., Tastes, ties, and time: A new social network dataset using facebook.com, *Social Networks*, 30, 4, 330-342, (2008)
- [65] Liaw, A.; Wiener, M., Classification and regression by randomforest, *R News*, 2, 3, 18-22, (2002)
- [66] Liaw, A. Wiener, M. (2012). randomForest: Breiman and Cutler's random forests for classification and regression. R package version 4.6-7.
- [67] Lima, E.; Mues, C.; Baesens, B., Domain knowledge integration in data mining using decision tables: case studies in churn prediction, *Journal of the Operational Research Society*, 60, 8, 1096-1106, (2009) · [Zbl 1168.90515](#)
- [68] Madden, G.; Savage, S. J.; Coble-Neal, G., Subscriber churn in the Australian ISP market, *Information Economics and Policy*, 11, 2, 195-207, (1999)
- [69] Martin-Barragan, B.; Lillo, R.; Romo, J., Interpretable support vector machines for functional data, *European Journal of Operational Research*, 232, 1, 146-155, (2014)
- [70] Matignon, R., *Neural Network Modeling Using Sas Enterprise Miner*, (2005), AuthorHouse
- [71] McCulloch, W.; Pitts, W., A logical calculus of the ideas immanent in nervous activity, *Bulletin of Mathematical Biophysics*, 5, 115-133, (1943) · [Zbl 0063.03860](#)
- [72] Meyer, D.; Dimitriadou, E.; Hornik, K.; Weingessel, A. Leisch, F. (2012). e1071: Misc functions of the department of statistics (e1071). R package version 1.6-1.
- [73] Migueis, V.; Van den Poel, D.; Camanho, A.; Falcao e Cunha, J., Modeling partial customer churn: on the value of first product-category purchase sequences, *Expert Systems with Applications*, 39, 12, 11250-11256, (2012)
- [74] Migueis, V. L.; Van den Poel, D.; Camanho, A. S.; Cunha, J. F.e., Predicting partial customer churn using Markov for discrimination for modeling first purchase sequences, *Advances in Data Analysis and Classification*, 6, 4, 337-353, (2012) ·

- [75] Mozer, M.; Wolniewicz, R.; Grimes, D.; Johnson, E.; Kaushansky, H., Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry, *IEEE Transactions on Neural Networks*, 11, 3, 690-696, (2000)
- [76] Nemenyi, P., Distribution-free multiple comparisons, (1963), Princeton University, Ph.D. thesis
- [77] Neslin, S. A.; Gupta, S.; Kamakura, W.; Junxiang, L.; Mason, C. H., Defection detection: measuring and understanding the predictive accuracy of customer churn models, *Journal of Marketing Research (JMR)*, 43, 2, 204-211, (2006)
- [78] Ngai, E.; Xiu, L.; Chau, D., Application of data mining techniques in customer relationship management: A literature review and classification, *Expert Systems with Applications*, 36, 2, 2592-2602, (2009)
- [79] Platt, J., Probabilistic outputs for support vector machines and comparison to regularized likelihood methods, *Advances in Large Margin Classifiers*, (2000), MIT Press Cambridge, MA
- [80] Prinzie, A.; Van den Poel, D., Incorporating sequential information into traditional classification models by using an element/position-sensitive SAM, *Decision Support Systems*, 42, 2, 508-526, (2006)
- [81] Prinzie, A.; Van den Poel, D., Random forests for multiclass classification: random multinomial logit, *Expert Systems with Applications*, 34, 3, 1721-1732, (2008)
- [82] Prinzie, A.; Van den Poel, D., Modeling complex longitudinal consumer behavior with dynamic Bayesian networks: an acquisition pattern analysis application, *Journal of Intelligent Information Systems*, 36, 3, 283-304, (2011)
- [83] Provost, F.; Fawcett, T.; Kohavi, R., The case against accuracy estimation for comparing induction algorithms, (Shavlik, J., Machine learning. proceedings of the fifteenth international conference on machine learning (ICML'98), (1998), Morgan Kaufmann Publishers Madison, WI, USA), 445-453
- [84] R Core Team (2013). stats: R statistical functions. R Foundation for Statistical Computing, Vienna, Austria, R package version 3.1.2.
- [85] Ripley, B., Pattern recognition and neural networks, (1996), Cambridge University Press · Zbl 0853.62046
- [86] Ripley, B. (2013). NNET: Feed-forward neural networks and multinomial log-linear models. R package version 7.3-8.
- [87] Rossi, P. E.; McCulloch, R. E.; Allenby, G. M., The value of purchase history data in target marketing, *Marketing Science*, 15, 4, 321-340, (1996)
- [88] Smith, K. A.; Willis, R. J.; Brooks, M., An analysis of customer retention and insurance claim patterns using data mining: A case study, *Journal of the Operational Research Society*, 51, 5, 532-541, (2000) · Zbl 1055.90575
- [89] Spackman, K. A., Maximum likelihood training of connectionist models: comparison with least squares back-propagation and logistic regression, *Proceedings of the annual symposium on computer application in medical care*, 285-289, (1991)
- [90] Thorleuchter, D.; Van den Poel, D., Predicting e-commerce company success by mining the text of its publicly-accessible website, *Expert Systems with Applications*, 39, 17, 13026-13034, (2012)
- [91] Thorleuchter, D.; Van den Poel, D.; Prinzie, A., Analyzing existing customers' websites to improve the customer acquisition process as well as the profitability prediction in b-to-b marketing, *Expert Systems with Applications*, 39, 3, 2597-2605, (2012)
- [92] Van den Poel, D.; Buckinx, W., Predicting online-purchasing behaviour, *European Journal of Operational Research*, 166, 2, 557-575, (2005) · Zbl 1064.90022
- [93] Van den Poel, D.; Lariviere, B., Customer attrition analysis for financial services using proportional hazard models, *European Journal of Operational Research*, 157, 1, 196-217, (2004) · Zbl 1106.91318
- [94] Venables, W. N.; Ripley, B. D., *Modern Applied Statistics with S*, (2002), Springer New York · Zbl 1006.62003
- [95] Venkatesh, K.; Ravi, V.; Prinzie, A.; Poel, D. V.d., Cash demand forecasting in ATMs by clustering and neural networks, *European Journal of Operational Research*, 232, 2, 383-392, (2014)
- [96] Verbeke, W.; Martens, D.; Mues, C.; Baesens, B., Building comprehensible customer churn prediction models with advanced rule induction techniques, *Expert Systems with Applications*, 38, 3, 2354-2364, (2011)
- [97] Verhaert, G. A.; Van den Poel, D., Improving campaign success rate by tailoring donation requests along the donor lifecycle, *Journal of Interactive Marketing*, 25, 1, 51-63, (2011)
- [98] Vert, J.-P.; Tsuda, K.; Schölkopf, B., A primer on kernel methods, (Schölkopf, B.; Tsuda, K.; Vert, J.-P., *Kernel methods in computational biology*, Computational Molecular Biology, (2004), MIT Press), 35-70
- [99] Weerahandi, S.; Moitra, S., Using survey data to predict adoption and switching for services, *Journal of Marketing Research*, 32, 1, 85-96, (1995)
- [100] Wei, C.-P.; Chiu, I.-T., Turning telecommunications call details to churn prediction: A data mining approach, *Expert Systems with Applications*, 23, 2, 103-112, (2002)
- [101] Xie, Y.; Li, X.; Ngai, E.; Ying, W., Customer churn prediction using improved balanced random forests, *Expert Systems with Applications*, 36, 3, 5445-5449, (2009)

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.