

Supplementary Table 5. Results Based on Modeling in CRISP-DM (N = 125)

Index	Author (Year)	Machine Learning Algorithms	Hyperparameter
A1	Marketta Hiissa et al. (2006)	LS-SVM	NI
A2	Manabu Nii et al. (2007)	SVM	Gaussian and Sigmoid kernels
A3	Laurence G. Moseley et al. (2008)	DT	NI
A4	Alexander Zlotnik et al. (2015)	SVR, M5P tree	SVM: (C = 1.0 epsilon = 1.0), M5P: (M = 4)
A5	Manabu Nii et al. (2016)	SVM	Grid Search, RBF kernel
A6	Robert Sherwin et al. (2017)	NI	NI
A7	Shinichiroh Yokota et al. (2017)	SVM	C (log2) = -10, Gamma(log2) = -2.5, ClassWeight = 627.29
A8	Jeungok Choi et al. (2018)	LR	L1 regularization.
A9	Stephen I. Gallant et al. (2018)	LR (L2)	NI
A10	Eliezer Bose et al. (2019)	GLM	glmnet $\alpha = 0.55$, $\lambda = 0.004$
A11	Gerald C. Gannod et al. (2019)	Apriori, LR	Minimum support 0.01, minimum confidence 0.8
A12	Steven G. Johnson et al. (2018)	Topic Modeling (TextRank), SVM	NI
A13	Zfania Tom Korach et al. (2019)	Topic modeling (LDA)	chunksize: 2,000, Passes:1, update_every:1, α (alpha): symmetric, eta: None, decay: 0.5, offset: 1, eval_every: 10, Iterations: 50, gamma_threshold: 0.001, minimum_probability: 0.01, random_state: None, minimum_phi_value: 0.01, per_word_topics: False
A14	Jae Yung Kwon et al. (2019)	LR, RF, XGBoost	NI
A15	Suzanne S. Sullivan et al. (2019)	DT, ANN, LR	NI
A16	Maxim Topaz et al. (2019)	RF	Default Settings
A17	Heather Brom et al. (2020)	LR, DT	k-fold cross-validation
A18	Roschelle L. Fritz et al. (2020)	RF, RT	trees = 100
A19	Christopher M. Horvat et al. (2020)	RF, LR, LR (L2), XGBoost, SVM, LR (L1), MARS	NI
A20	Renjie Hu et al. (2020)	ELMs+SOMs	NI
A21	Mireia Ladios-Martin et al. (2020)	LR	Iterative elimination of low-importance variables and tuning through permutation tests.
A22	Soo-Kyoung Lee et al. (2020)	RF, LR, SVM	trees = 100
A23	Chen Liang et al. (2020)	NB, SVM	NI
A24	David S. Lindberg et al. (2020)	DT, Bagging, RF, AdaBoost	trees = 500
A25	Jung In Park et al. (2020)	DT, LR, SVM	NI
A26	Maxim Topaz et al. (2020)	NB, DT, RF	Default Settings
A27	Dana M. Womack et al. (2020)	ET, GBM, RF, LR, SVM	Recursive feature elimination with cross-validation
A28	Ran An et al. (2021)	K-means, PLS Regression	PLS regression used cross-validation
A29	Linyan Chen et al. (2021)	SVM	NI
A30	Aaron Conway et al. (2021)	RF, LR, LR (L1), LR (L2), XGBoost	NI
A31	Alberto Garcés-Jiménez et al. (2021)	NB, RF, FC	Adjustments to the weights of the variables
A32	Li Hannaford et al. (2021)	RF, LR, DT, XGBoost, SVM, NB, KNN, ANN, Stacked Model	k-fold cross-validation
A33	Farinaz Havaei et al. (2021)	RF	10-fold cross-validation
A34	Elizabeth P. Howard et al. (2021)	Multiple LR	NI
A35	Mingyue Hu al. (2021)	RF, XGBoost, NB, LR	k-fold cross-validation
A36	Oleksandr Ivanov et al. (2021)	XGBoost	NI
A37	Liuqi Jin et al. (2021)	RF, KNN, SVM, DT	Particle Swarm Optimization
A38	Jisu Kim et al. (2021)	RF	NI
A39	Soo-Kyoung Lee et al. (2021)	RF, LR, SVM	NI

(Continued on the next page)

Supplementary Table 5. Continued

Index	Author (Year)	Machine Learning Algorithms	Hyperparameter
A40	Chia-Hui Liu et al. (2021)	DT, BN, SVM, RF, Bagging, AdaBoost	Default Settings
A41	Tamara G. R. Macieira et al. (2021)	RF	Default Settings
A42	Takuro Nagata et al. (2021)	SVM, RF	Default Settings
A43	Gojiro Nakagami et al. (2021)	LR, RF, SVM, XGBoost	Grid Search with 5-fold cross-validation
A44	Wenyu Song et al. (2021)	LR, SVM, RF, ANN	NI
A45	Rumei Yang et al. (2021)	RF, LR	Grid Search with cross-validation
A46	Huaqiong Zhou et al. (2021)	LR, RF, EN, GBM	Grid Search
A47	Yanhong Dong et al. (2022)	LightGBM, GBM, RF	Default Settings
A48	Farinaz Havaei et al. (2022)	RF	NI
A49	Tingting Hu et al. (2022)	LR, NB, SVM, AdaBoost	Grid Search With k-fold cross-validation
A50	Shuai Jin et al. (2022)	Linear Discriminant Analysis, LR, CT, RF, SVM	Grid Search with 10-fold cross-validation
A51	Mireia Ladios-Martin et al. (2022)	PCA, Averaged Perceptron, Bayes Point Machine, Boosted DT, Decision Forest, Decision Jungle, Deep SVM, LR, ANN, SVM	NI
A52	Young Ji Lee et al. (2022)	LR	NI
A53	Anup Kumar Mishra et al. (2022)	LR, DT, KNN, SVM, RF	Grid Search
A54	Kyoung Ja Moon et al. (2022)	C45, CBA, MCAR, LEM2, LR, ANN, SVM, RF	NI
A55	Nikhil Padhye et al. (2022)	LR (L1), ANN	NI
A56	Dongni Qian et al. (2022)	RF, LR	NI
A57	Javier Rojo et al. (2022)	Linear Regression	NI
A58	Jiyoun Song et al. (2022)	LR, RF, BayesianNetwork, SVM, NB	Optimization
A59	Tobias R. Spiller et al. (2022)	LR, RF, SVM, GBM, ANN	3 Repeats of 5-Fold Cross-Validation
A60	Katie Walker et al. (2022)	RF, LR, Elastic Net Regression	Optimization
A61	Melyana Nurul Widyawati et al. (2022)	DT, KNN, RF, ANN, AdaBoost, GNB, Bagging, ET, GBM, Stacking	NI
A62	Jie Xu et al. (2022)	LR, DT, RF	5-fold cross-validation
A63	Olga Yakusheva et al. (2022)	RF	10-fold cross-validation
A64	Ayla İrem Aydın et al. (2023)	Linear Regression	NI
A65	Rui CHEN et al. (2023)	KNN, SVM, NB, DT, RF, LR	NI
A66	Ya-Huei Chen et al. (2023)	XGBoost, NB, GLM, LR, DT, RF, ANN	Bayesian Optimization
A67	Pei-Yu Dai et al. (2023)	LR, RF, XGBoost, LightGBM	k-fold cross-validation
A68	Odai Y. Dweekat et al. (2023)	SVM	Default Settings
A69	Juliet Edgcomb et al. (2023)	RF, LR (L1)	NI
A70	Ajeet Gajra et al. (2023)	NI	NI
A71	Farinaz Havaei et al. (2023)	RF	NI
A72	Sharon Hewner et al. (2023)	K-means	NI
A73	Sunho Im et al. (2023)	XGBoost, RF, LR	k-fold cross-validation
A74	Junglyun Kim et al. (2023)	DT, RF, LR	10-fold cross-validation
A75	Seong-Kwang Kim et al. (2023)	RF, LR, DT	Grid Search with 5-fold cross-validation
A76	Hyungbok Lee et al. (2023)	LR, RF, NB, DT, SVM	10-fold cross-validation
A77	Hyungbok Lee et al. (2023)	LR, DT, NB, RF	NI
A78	Lin-Lin Lee et al. (2023)	K-means	NI

(Continued on the next page)

Supplementary Table 5. Continued

Index	Author (Year)	Machine Learning Algorithms	Hyperparameter
A79	Lingjuan Li et al. (2023)	LR, LR (L1), SVM, Linear Discriminant Analysis, RF, GBM, Ensemble	NI
A80	Pei-Hung Liao et al. (2023)	RF	NI
A81	Sarah R. Martha et al. (2023)	RF	trees = 5000
A82	Aruna Jothi Shanmugam et al. (2023)	SVM, RF, DT, Naïve Bayes, LR	NI
A83	Araceli Rodríguez Vico et al. (2023)	DT, Apriori	NI
A84	Zeping Yan et al. (2023)	DT, KNN, SVM, RF, LR	10-fold cross-validation
A85	Metin Yildiz et al. (2023)	KNN, SVM, ANN, RF, XGBoost, DT, GLMNET	10-fold cross-validation
A86	Ying Zhou et al. (2023)	RF, LR, SVM, KNN, NB, GBM, XGBoost	Optimization
A87	Maryam Zolnoori et al. (2023)	SVM, LR, RF, ET, AdaBoost, XGBoost	NI
A88	Young-Taek Park et al. (2024)	Linear Regressor, RF, GBM, XGBoost	Grid Search with 5-fold cross-validation
A89	Cynthia ABI KHALIL et al. (2024)	RF, DT, XGBoost, LR	5-fold cross-validation
A90	Yaser ALQARRAIN et al. (2024)	K-means	NI
A91	Komal Aryal et al. (2024)	LR, LR (L1), CART, RF, GBM	Grid Search with 10-fold cross-validation
A92	Ranjana Chavan et al. (2024)	NB, SVM, CN2 rule inducer, DT, ANN, GBM, SGD	Grid Search With k-fold cross-validation
A93	Xiaomei Chen et al. (2024)	DT, RF, GBM, XGBoost	Grid Search
A94	Colum Crowe et al. (2024)	LR, RF	trees = 100, depth = 2
A95	Tian Dai et al. (2024)	LR, SVM, KNN, RF, LightGBM, XGBoost	5-fold cross-validation
A96	Martha Duarte et al. (2024)	RF, ET, Bagging, XGBoost	5-fold cross-validation
A97	Yu-Fang Guo et al. (2024)	LR, SVM, RF, GBM	10-fold cross-validation
A98	Rui Jin et al. (2024)	LR, Linear Discriminant Analysis, RF, KNN, SVM	NI
A99	Arisa Kawashima et al. (2024)	XGBoost	Grid Search
A100	Yeonju Kim et al. (2024)	LR, NB, SVM, RF	10-fold cross-validation
A101	Ju Hee Lee et al. (2024)	GBM, LR, DT	NI
A102	Pin-Chieh Lee et al. (2024)	LR, ANN, BN, Tree base rule	NI
A103	Renee C. B. Manworren et al. (2024)	LR, RF, SVM	NI
A104	Ninon Girardon da Rosa et al. (2024)	RF	trees = 1,000, depth = 6
A105	Jihye Kim Scroggins et al. (2024)	XGBoost, SVM, LR, ET, RF	NI
A106	Lu Shao et al. (2024)	LR, GBM, XGBoost, RF, SVM, ANN, DT	Grid Search with 10-fold cross-validation
A107	Madeleine Stanik et al. (2024)	LR, RF, XGBoost, ANN	Grid Search With k-fold cross-validation
A108	Imam Tahyudin et al. (2024)	XGBoost, CatBoost, Extra Trees, DT, RF	Optimization
A109	Metin Yıldız et al. (2024)	KNN, SVM, ANN, RF, XGBoost, DT	10-fold cross-validation
A110	Cheng Yu et al. (2024)	XGBoost	5-fold cross-validation
A111	Wei Zhang et al. (2024)	RF, SVM, XGBoost, LR	Grid Search
A112	Maryam Zolnoori et al. (2024)	LR, RF, Extra Trees, AdaBoost, GBM, XGBoost, SVM	NI

(Continued on the next page)

Supplementary Table 5. Continued

Index	Author (Year)	Machine Learning Algorithms	Hyperparameter
A113	Yunping Zhang et al. (2022)	K-means	NI
A114	Goran Erfani et al. (2024)	K-means	Maximum 50 iterations, convergence criterion set at 0
A115	Tae Youn Kim et al. (2006)	LR, DT	Cut-off optimization for sensitivity/specificity balance
A116	In Sook Cho et al. (2011)	BN, LR	Cut-off points were tested
A117	Yoko Setoguchi et al. (2016)	DT	NI
A118	Mikyung Moon et al. (2017)	DT	NI
A119	Pacharmon Kaewprag et al. (2017)	BN	Evaluated combinations of search algorithms, scoring functions, and structure initializations
A120	Xiaohong Deng et al. (2017)	DT, LR	NI
A121	Hong-Lin Chen et al. (2018)	ANN	4 inputs, 1 bias neuron in input layer, 8 hidden neurons, 1 bias neuron in hidden layer, 1 output neuron
A122	Hsiu-Lan Li et al. (2019)	DT, ANN, SVM, LR	Trial-and-error method
A123	Seul Ki Park (2020)	LR	NI
A124	Sookyung Hyun et al. (2021)	LR	NI
A125	Ji-Yu CAI et al. (2021)	XGBoost	NI

AdaBoost: Adaptive Boosting; ANN: Artificial Neural Network; Apriori: Apriori Algorithm (Association Rule Mining); Bagging: Bootstrap Aggregating; BN: Bayesian Network; C45: C4.5 Decision Tree Algorithm; CART: Classification and Regression Tree; CatBoost: Categorical Boosting; CBA: Classification Based on Associations; CN2: CN2 Rule Inducer Algorithm; CT: Classification Tree; DT: Decision Tree; ELMS: Extreme Learning Machines; EN: Elastic Net Regression; ET: Extra Trees; FC: Filtered Classifier; GBM: Gradient Boosting Machine; GLM: Generalized Linear Model; GLMNET: Lasso and Elastic-Net Regularized Generalized Linear Models; GNB: Gaussian Naive Bayes; KNN: k-Nearest Neighbors; LDA: Latent Dirichlet Allocation (Topic Modeling); LEM2: Learning from Examples Module 2; LightGBM: Light Gradient Boosting Machine; LR: Logistic Regression; LR (L1): Logistic Regression with L1 Regularization; LR (L2): Logistic Regression with L2 Regularization; M5P: M5P Regression Tree (Model Tree Algorithm); MARS: Multivariate Adaptive Regression Splines; MCAR: Modified Class Association Rules; NB: Naive Bayes; NI: No Information; PCA: Principal Component Analysis; PLS: Partial Least Squares Regression; RF: Random Forest; RT: Regression Tree; SGD: Stochastic Gradient Descent; SOMs: Self-Organizing Maps; SVM: Support Vector Machine; SVR: Support Vector Regression; XGBoost: eXtreme Gradient Boosting.