|  |  |
| --- | --- |
| **Adult Data Set**  *Download*: [Data Folder](http://archive.ics.uci.edu/ml/machine-learning-databases/adult/), [Data Set Description](http://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.names)  **Abstract**: Predict whether income exceeds $50K/yr based on census data. Also known as "Census Income" dataset. | http://archive.ics.uci.edu/ml/assets/MLimages/Large2.jpg |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Data Set Characteristics:** | Multivariate | **Number of Instances:** | 48842 | **Area:** | Social |
| **Attribute Characteristics:** | Categorical, Integer | **Number of Attributes:** | 14 | **Date Donated** | 1996-05-01 |
| **Associated Tasks:** | Classification | **Missing Values?** | Yes | **Number of Web Hits:** | 801410 |

**Source:**

Donor:   
  
Ronny Kohavi and Barry Becker   
Data Mining and Visualization   
Silicon Graphics.   
e-mail: ronnyk '@' live.com for questions.

**Data Set Information:**

Extraction was done by Barry Becker from the 1994 Census database. A set of reasonably clean records was extracted using the following conditions: ((AAGE>16) && (AGI>100) && (AFNLWGT>1)&& (HRSWK>0))   
  
Prediction task is to determine whether a person makes over 50K a year.

**Attribute Information:**

Listing of attributes:   
  
>50K, <=50K.   
  
age: continuous.   
workclass: Private, Self-emp-not-inc, Self-emp-inc, Federal-gov, Local-gov, State-gov, Without-pay, Never-worked.   
fnlwgt: continuous.   
education: Bachelors, Some-college, 11th, HS-grad, Prof-school, Assoc-acdm, Assoc-voc, 9th, 7th-8th, 12th, Masters, 1st-4th, 10th, Doctorate, 5th-6th, Preschool.   
education-num: continuous.   
marital-status: Married-civ-spouse, Divorced, Never-married, Separated, Widowed, Married-spouse-absent, Married-AF-spouse.   
occupation: Tech-support, Craft-repair, Other-service, Sales, Exec-managerial, Prof-specialty, Handlers-cleaners, Machine-op-inspct, Adm-clerical, Farming-fishing, Transport-moving, Priv-house-serv, Protective-serv, Armed-Forces.   
relationship: Wife, Own-child, Husband, Not-in-family, Other-relative, Unmarried.   
race: White, Asian-Pac-Islander, Amer-Indian-Eskimo, Other, Black.   
sex: Female, Male.   
capital-gain: continuous.   
capital-loss: continuous.   
hours-per-week: continuous.   
native-country: United-States, Cambodia, England, Puerto-Rico, Canada, Germany, Outlying-US(Guam-USVI-etc), India, Japan, Greece, South, China, Cuba, Iran, Honduras, Philippines, Italy, Poland, Jamaica, Vietnam, Mexico, Portugal, Ireland, France, Dominican-Republic, Laos, Ecuador, Taiwan, Haiti, Columbia, Hungary, Guatemala, Nicaragua, Scotland, Thailand, Yugoslavia, El-Salvador, Trinadad&Tobago, Peru, Hong, Holand-Netherlands.

**Relevant Papers:**

Ron Kohavi, "Scaling Up the Accuracy of Naive-Bayes Classifiers: a Decision-Tree Hybrid", Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, 1996   
[[Web Link]](http://robotics.stanford.edu/~ronnyk/nbtree.pdf)

**Papers That Cite This Data Set1:**

http://archive.ics.uci.edu/ml/assets/rexa.jpg

Rakesh Agrawal and Ramakrishnan ikant and Dilys Thomas. [Privacy Preserving OLAP](http://rexa.info/paper/686ccd571de3e92d89394f207348bd5a669558d2). SIGMOD Conference. 2005. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#686ccd571de3e92d89394f207348bd5a669558d2)].  
  
Rich Caruana and Alexandru Niculescu-Mizil. [An Empirical Evaluation of Supervised Learning for ROC Area](http://rexa.info/paper/97c489fe4c5d81f7fde00a62d3a5d043ec9ddcb6). ROCAI. 2004. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#97c489fe4c5d81f7fde00a62d3a5d043ec9ddcb6)].  
  
Rich Caruana and Alexandru Niculescu-Mizil and Geoff Crew and Alex Ksikes. [Ensemble selection from libraries of models](http://rexa.info/paper/b2c60bdf066b03909792bd5d64891f3607b948c8). ICML. 2004. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#b2c60bdf066b03909792bd5d64891f3607b948c8)].  
  
Bianca Zadrozny. [Learning and evaluating classifiers under sample selection bias](http://rexa.info/paper/1d416aa15df505a2052d91f04ce93d8e3f2bfa7e). ICML. 2004. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#1d416aa15df505a2052d91f04ce93d8e3f2bfa7e)].  
  
Wei-Chun Kao and Kai-Min Chung and Lucas Assun and Chih-Jen Lin. [Decomposition Methods for Linear Support Vector Machines](http://rexa.info/paper/1383360e8bb2cfd7a98219f867869a9f6d7e0db0). Neural Computation, 16. 2004. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#1383360e8bb2cfd7a98219f867869a9f6d7e0db0)].  
  
Saharon Rosset. [Model selection via the AUC](http://rexa.info/paper/92270daca0b7d369e25eb444e918bcb74b18d682). ICML. 2004. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#92270daca0b7d369e25eb444e918bcb74b18d682)].  
  
I. Yoncaci. [Maximum a Posteriori Tree Augmented Naive Bayes Classifiers](http://rexa.info/paper/df1dc2bd013ef17735eb85d2de85701f833db3ff). O EN INTEL.LIG ` ENCIA ARTIFICIAL CSIC. 2003. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#df1dc2bd013ef17735eb85d2de85701f833db3ff)].  
  
Christopher R. Palmer and Christos Faloutsos. [Electricity Based External Similarity of Categorical Attributes](http://rexa.info/paper/5e9e56b4525a16e039d75d04d32477b118e36b0d). PAKDD. 2003. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#5e9e56b4525a16e039d75d04d32477b118e36b0d)].  
  
S. Sathiya Keerthi and Chih-Jen Lin. [Asymptotic Behaviors of Support Vector Machines with Gaussian Kernel](http://rexa.info/paper/679efdb662f70091c573c74a40aadab38fc0d868). Neural Computation, 15. 2003. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#679efdb662f70091c573c74a40aadab38fc0d868)].  
  
Thomas Serafini and G. Zanghirati and Del Zanna and T. Serafini and Gaetano Zanghirati and Luca Zanni. [DIPARTIMENTO DI MATEMATICA](http://rexa.info/paper/a702c0617d7187cfc988819d79f0eb2a42ba3f19). Gradient Projection Methods for. 2003. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#a702c0617d7187cfc988819d79f0eb2a42ba3f19)].  
  
Bart Hamers and J. A. K Suykens. [Coupled Transductive Ensemble Learning of Kernel Models](http://rexa.info/paper/74b68cad54c62b62b7ad64911291e2e0b0620dd8). Bart De Moor. 2003. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#74b68cad54c62b62b7ad64911291e2e0b0620dd8)].  
  
Andrew W. Moore and Weng-Keen Wong. [Optimal Reinsertion: A New Search Operator for Accelerated and More Accurate Bayesian Network Structure Learning](http://rexa.info/paper/13c8e4e9474a8fbd7bd8a66695b62f218189ac6f). ICML. 2003. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#13c8e4e9474a8fbd7bd8a66695b62f218189ac6f)].  
  
Alexander J. Smola and Vishy Vishwanathan and Eleazar Eskin. [Laplace Propagation](http://rexa.info/paper/0767f0171b0f74f96978a41f3e947ea91cc9dbd3). NIPS. 2003. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#0767f0171b0f74f96978a41f3e947ea91cc9dbd3)].  
  
Bianca Zadrozny and Charles Elkan. [Transforming classifier scores into accurate multiclass probability estimates](http://rexa.info/paper/4284c9cb6236847cd246f69cfb8e4209c107d18f). KDD. 2002. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#4284c9cb6236847cd246f69cfb8e4209c107d18f)].  
  
Nitesh V. Chawla and Kevin W. Bowyer and Lawrence O. Hall and W. Philip Kegelmeyer. [SMOTE: Synthetic Minority Over-sampling Technique](http://rexa.info/paper/4dbb0e14d9556fd1099e129462d0bedcf35bd82b). J. Artif. Intell. Res. (JAIR, 16. 2002. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#4dbb0e14d9556fd1099e129462d0bedcf35bd82b)].  
  
S. Sathiya Keerthi and Kaibo Duan and Shirish Krishnaj Shevade and Aun Neow Poo. [A Fast Dual Algorithm for Kernel Logistic Regression](http://rexa.info/paper/6a6b4ca13137132e00185012fd7c4667ff964769). ICML. 2002. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#6a6b4ca13137132e00185012fd7c4667ff964769)].  
  
Ramesh Natarajan and Edwin P D Pednault. [Segmented Regression Estimators for Massive Data Sets](http://rexa.info/paper/81ff15994633ecdb2b9fd5a6d11d985c2b215442). SDM. 2002. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#81ff15994633ecdb2b9fd5a6d11d985c2b215442)].  
  
Jie Cheng and Russell Greiner. [Learning Bayesian Belief Network Classifiers: Algorithms and System](http://rexa.info/paper/beeb203c082359f4e141e1767a14f09449a5a717). Canadian Conference on AI. 2001. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#beeb203c082359f4e141e1767a14f09449a5a717)].  
  
Zhiyuan Chen and Johannes Gehrke and Flip Korn. [Query Optimization In Compressed Database Systems](http://rexa.info/paper/6bacbb21f7a1d6dee581f756cf326e0932446698). SIGMOD Conference. 2001. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#6bacbb21f7a1d6dee581f756cf326e0932446698)].  
  
Stephen D. Bay. [Multivariate Discretization for Set Mining](http://rexa.info/paper/0527f760f533e2a7a827710a924d6987bfc9a06e). Knowl. Inf. Syst, 3. 2001. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#0527f760f533e2a7a827710a924d6987bfc9a06e)].  
  
Bernhard Pfahringer and Geoffrey Holmes and Richard Kirkby. [Optimizing the Induction of Alternating Decision Trees](http://rexa.info/paper/510872c81d0a05f14c583cf3e8245b6c88d43db6). PAKDD. 2001. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#510872c81d0a05f14c583cf3e8245b6c88d43db6)].  
  
Stephen D. Bay and Michael J. Pazzani. [Detecting Group Differences: Mining Contrast Sets](http://rexa.info/paper/8931d2a4a8256ea88abea3ea3ac820fd421ce0b1). Data Min. Knowl. Discov, 5. 2001. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#8931d2a4a8256ea88abea3ea3ac820fd421ce0b1)].  
  
Dmitry Pavlov and Jianchang Mao and Byron Dom. [Scaling-Up Support Vector Machines Using Boosting Algorithm](http://rexa.info/paper/d6a42d10713c10dbc3efa4b7de53fb7f807caa0b). ICPR. 2000. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#d6a42d10713c10dbc3efa4b7de53fb7f807caa0b)].  
  
Gary M. Weiss and Haym Hirsh. [A Quantitative Study of Small Disjuncts: Experiments and Results](http://rexa.info/paper/55a7f45a6b42ceb9097c7260032694bcbd8a2fd2). Department of Computer Science Rutgers University. 2000. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#55a7f45a6b42ceb9097c7260032694bcbd8a2fd2)].  
  
Dmitry Pavlov and Darya Chudova and Padhraic Smyth. [Towards scalable support vector machines using squashing](http://rexa.info/paper/5300c71ae945611a1d091404fc0461a48738e151). KDD. 2000. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#5300c71ae945611a1d091404fc0461a48738e151)].  
  
Kristin P. Bennett and Ayhan Demiriz and John Shawe-Taylor. [A Column Generation Algorithm For Boosting](http://rexa.info/paper/94cf84d5a454c97d326efdb24b2b96e0eac25c33). ICML. 2000. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#94cf84d5a454c97d326efdb24b2b96e0eac25c33)].  
  
Petri Kontkanen and Jussi Lahtinen and Petri Myllymaki and Tomi Silander and Henry Tirri. [Proceedings of Pre- and Post-processing in Machine Learning and Data Mining: Theoretical Aspects and Applications, a workshop within Machine Learning and Applications](http://rexa.info/paper/819bb37e80fcb500fd159702e1dc2c0fe98f38ac). Complex Systems Computation Group (CoSCo). 1999. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#819bb37e80fcb500fd159702e1dc2c0fe98f38ac)].  
  
Jie Cheng and Russell Greiner. [Comparing Bayesian Network Classifiers](http://rexa.info/paper/47354ca48da5014e0a8f5e4da7f3a7e9aaa6e9e5). UAI. 1999. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#47354ca48da5014e0a8f5e4da7f3a7e9aaa6e9e5)].  
  
Yk Huhtala and Juha Kärkkäinen and Pasi Porkka and Hannu Toivonen. [Efficient Discovery of Functional and Approximate Dependencies Using Partitions](http://rexa.info/paper/ac8fe867e1d16d4d09f9bd759ba46699055c7ca6). ICDE. 1998. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#ac8fe867e1d16d4d09f9bd759ba46699055c7ca6)].  
  
John C. Platt. [Using Analytic QP and Sparseness to Speed Training of Support Vector Machines](http://rexa.info/paper/46a760d333f90e12eb326d2b17c90aa70c01d1df). NIPS. 1998. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#46a760d333f90e12eb326d2b17c90aa70c01d1df)].  
  
Ron Kohavi. [Scaling Up the Accuracy of Naive-Bayes Classifiers: A Decision-Tree Hybrid](http://rexa.info/paper/bf6cec50b7f7d48d105c8c649210cc3a42d3d71e). KDD. 1996. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#bf6cec50b7f7d48d105c8c649210cc3a42d3d71e)].  
  
William W. Cohen and Yoram Singer. [A Simple, Fast, and Effective Rule Learner](http://rexa.info/paper/0ded09c3d904c9f8eeef9ed0bc3445c2b44a6008). AT&T Labs--Research Shannon Laboratory. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#0ded09c3d904c9f8eeef9ed0bc3445c2b44a6008)].  
  
Haixun Wang and Philip S. Yu. [SSDT-NN: A Subspace-Splitting Decision Tree Classifier with Application to Target Selection](http://rexa.info/paper/c7d3013a5143009f65d4303e65d8ecfcc892fa2e). IBM T. J. Watson Research Center. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#c7d3013a5143009f65d4303e65d8ecfcc892fa2e)].  
  
S. V. N Vishwanathan and Alexander J. Smola and M. Narasimha Murty. [considerably faster than competing methods such as Sequential Minimal Optimization or the Nearest Point Algorithm](http://rexa.info/paper/cefc3da70b90693b8046b574392c86d2db1f196a). Machine Learning Program, National ICT for Australia. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#cefc3da70b90693b8046b574392c86d2db1f196a)].  
  
Grigorios Tsoumakas and Ioannis P. Vlahavas. [Fuzzy Meta-Learning: Preliminary Results](http://rexa.info/paper/7de1dca96d7e6789dadc875a7f1b070f9d65fff5). Greek Secretariat for Research and Technology. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#7de1dca96d7e6789dadc875a7f1b070f9d65fff5)].  
  
Josep Roure Alcobe. [Incremental Hill-Climbing Search Applied to Bayesian Network Structure Learning](http://rexa.info/paper/59e7c62dd41ce2ed1741bafac723745eb7b80b65). Escola Universitria Politcnica de Mataro. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#59e7c62dd41ce2ed1741bafac723745eb7b80b65)].  
  
Ayhan Demiriz and Kristin P. Bennett and John Shawe and I. Nouretdinov V.. [Linear Programming Boosting via Column Generation](http://rexa.info/paper/fc0a66d3a7336b6eabad919a7389c68cc37f2564). Dept. of Decision Sciences and Eng. Systems, Rensselaer Polytechnic Institute. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#fc0a66d3a7336b6eabad919a7389c68cc37f2564)].  
  
Chris Giannella and Bassem Sayrafi. [An Information Theoretic Histogram for Single Dimensional Selectivity Estimation](http://rexa.info/paper/48dc427a3536d7bef34e910efc1f34171bf98729). Department of Computer Science, Indiana University Bloomington. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#48dc427a3536d7bef34e910efc1f34171bf98729)].  
  
Rong-En Fan and P. -H Chen and C. -J Lin. [Working Set Selection Using the Second Order Information for Training SVM](http://rexa.info/paper/da329267bf8880c2becb15eae121a5b002347349). Department of Computer Science and Information Engineering National Taiwan University. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#da329267bf8880c2becb15eae121a5b002347349)].  
  
Petri Kontkanen and Jussi Lahtinen and Petri Myllymaki and Tomi Silander and Henry Tirri. [USING BAYESIAN NETWORKS FOR VISUALIZING HIGH-DIMENSIONAL DATA](http://rexa.info/paper/da53589d7ad993c65a1857401ea6f7b0b025f41a). Complex Systems Computation Group (CoSCo). [[View Context](http://archive.ics.uci.edu/ml/support/Adult#da53589d7ad993c65a1857401ea6f7b0b025f41a)].  
  
Ahmed Hussain Khan and Intensive Care. [Multiplier-Free Feedforward Networks](http://rexa.info/paper/ef65a20ee05551cc411c8acd0c5af4796dc1a39e). 174. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#ef65a20ee05551cc411c8acd0c5af4796dc1a39e)].  
  
Luc Hoegaerts and J. A. K Suykens and J. Vandewalle and Bart De Moor. [Subset Based Least Squares Subspace Regression in RKHS](http://rexa.info/paper/e3cb1b3c129bc2be18036ad91073724543894185). Katholieke Universiteit Leuven Department of Electrical Engineering, ESAT-SCD-SISTA. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#e3cb1b3c129bc2be18036ad91073724543894185)].  
  
David R. Musicant and Alexander Feinberg. [Active Set Support Vector Regression](http://rexa.info/paper/8e003bfae7c07cdeeb0ae5f251d545873364dd1e). [[View Context](http://archive.ics.uci.edu/ml/support/Adult#8e003bfae7c07cdeeb0ae5f251d545873364dd1e)].  
  
Luc Hoegaerts and J. A. K Suykens and J. Vandewalle and Bart De Moor. [Primal Space Sparse Kernel Partial Least Squares Regression for Large Scale Problems Special Session paper](http://rexa.info/paper/c26b03be3bf94e423d30b545d09efd83310e97bd) . Katholieke Universiteit Leuven Department of Electrical Engineering, ESAT-SCD-SISTA. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#c26b03be3bf94e423d30b545d09efd83310e97bd)].  
  
Kuan-ming Lin and Chih-Jen Lin. [A Study on Reduced Support Vector Machines](http://rexa.info/paper/ce160518d6a429585aeeb3f7a784c5dfb124b669). Department of Computer Science and Information Engineering National Taiwan University. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#ce160518d6a429585aeeb3f7a784c5dfb124b669)].  
  
Luca Zanni. [An Improved Gradient Projection-based Decomposition Technique for Support Vector Machines](http://rexa.info/paper/0e6263639bcf297001939fd27576b20c6b78f68b). Dipartimento di Matematica, Universitdi Modena e Reggio Emilia. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#0e6263639bcf297001939fd27576b20c6b78f68b)].  
  
Jeff G. Schneider and Andrew W. Moore. [Active Learning in Discrete Input Spaces](http://rexa.info/paper/8ea9716e5199404825766638c2cc07ec93be5236). School of Computer Science Carnegie Mellon University. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#8ea9716e5199404825766638c2cc07ec93be5236)].  
  
Omid Madani and David M. Pennock and Gary William Flake. [Co-Validation: Using Model Disagreement to Validate Classification Algorithms](http://rexa.info/paper/03a71aaf988c71c8022be08734da8e376f7fe037). Yahoo! Research Labs. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#03a71aaf988c71c8022be08734da8e376f7fe037)].  
  
Ron Kohavi and Barry G. Becker and Dan Sommerfield. [Improving Simple Bayes](http://rexa.info/paper/4c8e8cf6857f1f1bc9b43679d241b096513ee6f2). Data Mining and Visualization Group Silicon Graphics, Inc. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#4c8e8cf6857f1f1bc9b43679d241b096513ee6f2)].  
  
Shi Zhong and Weiyu Tang and Taghi M. Khoshgoftaar. [Boosted Noise Filters for Identifying Mislabeled Data](http://rexa.info/paper/e2e72927eb590e2b7daf9095e42d6ed43ce21e68). Department of Computer Science and Engineering Florida Atlantic University. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#e2e72927eb590e2b7daf9095e42d6ed43ce21e68)].  
  
David R. Musicant. [DATA MINING VIA MATHEMATICAL PROGRAMMING AND MACHINE LEARNING](http://rexa.info/paper/f1395e1da4a724219d6cc414e48969140355bebb). Doctor of Philosophy (Computer Sciences) UNIVERSITY. [[View Context](http://archive.ics.uci.edu/ml/support/Adult#f1395e1da4a724219d6cc414e48969140355bebb)].