

This data was extracted from the census bureau database found at
<http://www.census.gov/ftp/pub/DES/www/welcome.html>
 Donor: Terran Lane and Ronny Kohavi
 Data Mining and Visualization
 Silicon Graphics.
 e-mail: terran@ecn.purdue.edu, ronnyk@sgi.com for questions.

The data was split into train/test in approximately 2/3, 1/3 proportions using MineSet's MIndUtil mineset-to-mlc.

Prediction task is to determine the income level for the person represented by the record. Incomes have been binned at the \$50K level to present a binary classification problem, much like the original UCI/ADULT database. The goal field of this data, however, was drawn from the "total person income" field rather than the "adjusted gross income" and may, therefore, behave differently than the original ADULT goal field.

More information detailing the meaning of the attributes can be found in <http://www.bls.census.gov/cps/cpsmain.htm>
 To make use of the data descriptions at this site, the following mappings to the Census Bureau's internal database column names will be needed:

age	AAGE
class of worker	ACLSWKR
industry code	ADTIND
occupation code	ADTOCC
adjusted gross income	AGI
education	AHGA
wage per hour	AHRSPAY
enrolled in edu inst last wk	AHSCOL
marital status	AMARITL
major industry code	AMJIND
major occupation code	AMJOCC
mace	ARACE
hispanic Origin	AREORGN
sex	ASEX
member of a labor union	AUNMEM
reason for unemployment	AUNTYPE
full or part time employment stat	AWKSTAT
capital gains	CAPGAIN
capital losses	CAPLOSS
divdends from stocks	DIVVAL
federal income tax liability	FEDTAX
tax filer status	FILESTAT
region of previous residence	GRINREG
state of previous residence	GRINST
detailed household and family stat	HHDFMX
detailed household summary in household	HHDREL
instance weight	MARSUPWT
migration code-change in msa	MIGMTR1
migration code-change in reg	MIGMTR3
migration code-move within reg	MIGMTR4
live in this house 1 year ago	MIGSAME
migration prev res in sunbelt	MIGSUN
num persons worked for employer	NOEMP
family members under 18	PARENT
total person earnings	PEARVAL
country of birth father	PEFNTVTY
country of birth mother	PEMNTVTY
country of birth self	PENATVTY
citizenship	PRCITSHP
total person income	PTOTVAL
own business or self employed	SEOTR

taxable income amount	TAXINC
fill inc questionnaire for veteran's admin	VETQVA
veterans benefits	VETYN
weeks worked in year	WKSWORK

Basic statistics for this data set:

Number of instances data = 199523

Duplicate or conflicting instances : 46716

Number of instances in test = 99762

Duplicate or conflicting instances : 20936

Class probabilities for income-projected.test file

Probability for the label '- 50000' : 93.80%

Probability for the label '50000+' : 6.20%

Majority accuracy: 93.80% on value - 50000

Number of attributes = 40 (continuous : 7 nominal : 33)

Information about .data file :

```

91 distinct values for attribute #0 (age) continuous
9 distinct values for attribute #1 (class of worker) nominal
52 distinct values for attribute #2 (detailed industry recode) nominal
47 distinct values for attribute #3 (detailed occupation recode) nominal
17 distinct values for attribute #4 (education) nominal
1240 distinct values for attribute #5 (wage per hour) continuous
3 distinct values for attribute #6 (enroll in edu inst last wk) nominal
7 distinct values for attribute #7 (marital stat) nominal
24 distinct values for attribute #8 (major industry code) nominal
15 distinct values for attribute #9 (major occupation code) nominal
5 distinct values for attribute #10 (race) nominal
10 distinct values for attribute #11 (hispanic origin) nominal
2 distinct values for attribute #12 (sex) nominal
3 distinct values for attribute #13 (member of a labor union) nominal
6 distinct values for attribute #14 (reason for unemployment) nominal
8 distinct values for attribute #15 (full or part time employment stat) nominal
132 distinct values for attribute #16 (capital gains) continuous
113 distinct values for attribute #17 (capital losses) continuous
1478 distinct values for attribute #18 (dividends from stocks) continuous
6 distinct values for attribute #19 (tax filer stat) nominal
6 distinct values for attribute #20 (region of previous residence) nominal
51 distinct values for attribute #21 (state of previous residence) nominal
38 distinct values for attribute #22 (detailed household and family stat) nominal
8 distinct values for attribute #23 (detailed household summary in household) nominal
10 distinct values for attribute #24 (migration code-change in msa) nominal
9 distinct values for attribute #25 (migration code-change in reg) nominal
10 distinct values for attribute #26 (migration code-move within reg) nominal
3 distinct values for attribute #27 (live in this house 1 year ago) nominal
4 distinct values for attribute #28 (migration prev res in sunbelt) nominal
7 distinct values for attribute #29 (num persons worked for employer) continuous
5 distinct values for attribute #30 (family members under 18) nominal
43 distinct values for attribute #31 (country of birth father) nominal
43 distinct values for attribute #32 (country of birth mother) nominal
43 distinct values for attribute #33 (country of birth self) nominal
5 distinct values for attribute #34 (citizenship) nominal
3 distinct values for attribute #35 (own business or self employed) nominal
3 distinct values for attribute #36 (fill inc questionnaire for veteran's admin) nominal
3 distinct values for attribute #37 (veterans benefits) nominal
53 distinct values for attribute #38 (weeks worked in year) continuous
2 distinct values for attribute #39 (year) nominal

```

Error rates:

C4.5	: 4.8%
C5.0	: 4.7%
C5.0 rules	: 4.7%
C5.0 boosting	: 4.6%
Naive-Bayes	: 23.2%

All commas and periods were changed to spaces
Colons were replaced with dashes.

The instance weight indicates the number of people in the population that each record represents due to stratified sampling.
To do real analysis and derive conclusions, this field must be used.
This attribute should *not* be used in the classifiers, so it is set to "ignore" in this file.

- 50000, 50000+.

age: continuous.

class of worker: Not in universe, Federal government, Local government, Never worked, Private, Self-employed-incorporated, Self-employed-not incorporated, State government, Without pay.

detailed industry recode: 0, 40, 44, 2, 43, 47, 48, 1, 11, 19, 24, 25, 32, 33, 34, 35, 36, 37, 38, 39, 4, 42, 45, 5, 15, 16, 22, 29, 31, 50, 14, 17, 18, 28, 3, 30, 41, 46, 51, 12, 13, 21, 23, 26, 6, 7, 9, 49, 27, 8, 10, 20.

detailed occupation recode: 0, 12, 31, 44, 19, 32, 10, 23, 26, 28, 29, 42, 40, 34, 14, 36, 38, 2, 20, 25, 37, 41, 27, 24, 30, 43, 33, 16, 45, 17, 35, 22, 18, 39, 3, 15, 13, 46, 8, 21, 9, 4, 6, 5, 1, 11, 7.

education: Children, 7th and 8th grade, 9th grade, 10th grade, High school graduate, 11th grade, 12th grade no diploma, 5th or 6th grade, Less than 1st grade, Bachelors degree(BA AB BS), 1st 2nd 3rd or 4th grade, Some college but no degree, Masters degree(MA MS MEng MEd MSW MBA), Associates degree-occup /vocational, Associates degree-academic program, Doctorate degree(PhD EdD), Prof school degree (MD DDS DVM LLB JD).

wage per hour: continuous.

enroll in edu inst last wk: Not in universe, High school, College or university.

marital stat: Never married, Married-civilian spouse present, Married-spouse absent, Separated, Divorced, Widowed, Married-A F spouse present.

major industry code: Not in universe or children, Entertainment, Social services, Agriculture, Education, Public administration, Manufacturing-durable goods, Manufacturing-nondurable goods, Wholesale trade, Retail trade, Finance insurance and real estate, Private household services, Business and repair services, Personal services except private HH, Construction, Medical except hospital, Other professional services, Transportation, Utilities and sanitary services, Mining, Communications, Hospital services, Forestry and fisheries, Armed Forces.

major occupation code: Not in universe, Professional specialty, Other service, Farming forestry and fishing, Sales, Adm support including clerical, Protective services, Handlers equip cleaners etc , Precision production craft & repair, Technicians and related support, Machine operators assmblrs & inspctrs, Transportation and material moving, Executive admin and managerial, Private household services, Armed Forces.

race: White, Black, Other, Amer Indian Aleut or Eskimo, Asian or Pacific Islander.

hispanic origin: Mexican (Mexicano), Mexican-American, Puerto Rican, Central or South American, All other, Other Spanish, Chicano, Cuban, Do not know, NA.

sex: Female, Male.

member of a labor union: Not in universe, No, Yes.

reason for unemployment: Not in universe, Re-entrant, Job loser - on layoff, New entrant, Job leaver, Other job loser.

full or part time employment stat: Children or Armed Forces, Full-time schedules, Unemployed part- time, Not in labor force, Unemployed full-time, PT for non-econ reasons usually FT, PT for econ reasons usually PT, PT for econ reasons usually FT.

capital gains: continuous.

capital losses: continuous.

dividends from stocks: continuous.

tax filer stat: Nonfiler, Joint one under 65 & one 65+, Joint both under 65, Single, Head of household, Joint both 65+.

region of previous residence: Not in universe, South, Northeast, West, Midwest, Abroad.

state of previous residence: Not in universe, Utah, Michigan, North Carolina, North Dakota, Virginia, Vermont, Wyoming, West Virginia, Pennsylvania, Abroad, Oregon, California, Iowa, Florida, Arkansas, Texas, South Carolina, Arizona, Indiana, Tennessee, Maine, Alaska, Ohio, Montana, Nebraska, Mississippi, District of Columbia, Minnesota, Illinois, Kentucky, Delaware, Colorado, Maryland, Wisconsin, New Hampshire, Nevada, New York, Georgia, Oklahoma,

New Mexico, South Dakota, Missouri, Kansas, Connecticut, Louisiana, Alabama, Massachusetts, Idaho, New Jersey.

detailed household and family stat: Child <18 never marr not in subfamily, Other Rel <18 never marr child of subfamily RP, Other Rel <18 never marr not in subfamily, Grandchild <18 never marr child of subfamily RP, Grandchild <18 never marr not in subfamily, Secondary individual, In group quarters, Child under 18 of RP of unrel subfamily, RP of unrelated subfamily, Spouse of householder, Householder, Other Rel <18 never married RP of subfamily, Grandchild <18 never marr RP of subfamily, Child <18 never marr RP of subfamily, Child <18 ever marr not in subfamily, Other Rel <18 ever marr RP of subfamily, Child <18 ever marr RP of subfamily, Nonfamily householder, Child <18 spouse of subfamily RP, Other Rel <18 spouse of subfamily RP, Other Rel <18 ever marr not in subfamily, Grandchild <18 ever marr not in subfamily, Child 18+ never marr Not in a subfamily, Grandchild 18+ never marr not in subfamily, Child 18+ ever marr RP of subfamily, Other Rel 18+ never marr not in subfamily, Child 18+ never marr RP of subfamily, Other Rel 18+ ever marr RP of subfamily, Other Rel 18+ never marr RP of subfamily, Other Rel 18+ spouse of subfamily RP, Other Rel 18+ ever marr not in subfamily, Child 18+ ever marr Not in a subfamily, Grandchild 18+ ever marr not in subfamily, Child 18+ spouse of subfamily RP, Spouse of RP of unrelated subfamily, Grandchild 18+ ever marr RP of subfamily, Grandchild 18+ never marr RP of subfamily, Grandchild 18+ spouse of subfamily RP.

detailed household summary in household: Child under 18 never married, Other relative of householder, Nonrelative of householder, Spouse of householder, Householder, Child under 18 ever married, Group Quarters- Secondary individual, Child 18 or older.

| instance weight: ignore.

instance weight: continuous.

migration code-change in msa: Not in universe, Nonmover, MSA to MSA, NonMSA to nonMSA, MSA to nonMSA, NonMSA to MSA, Abroad to MSA, Not identifiable, Abroad to nonMSA.

migration code-change in reg: Not in universe, Nonmover, Same county, Different county same state, Different state same division, Abroad, Different region, Different division same region.

migration code-move within reg: Not in universe, Nonmover, Same county, Different county same state, Different state in West, Abroad, Different state in Midwest, Different state in South, Different state in Northeast.

live in this house 1 year ago: Not in universe under 1 year old, Yes, No.

migration prev res in sunbelt: Not in universe, Yes, No.

num persons worked for employer: continuous.

family members under 18: Both parents present, Neither parent present, Mother only present, Father only present, Not in universe.

country of birth father: Mexico, United-States, Puerto-Rico, Dominican-Republic, Jamaica, Cuba, Portugal, Nicaragua, Peru, Ecuador, Guatemala, Philippines, Canada, Columbia, El-Salvador, Japan, England, Trinidad&Tobago, Honduras, Germany, Taiwan, Outlying-U S (Guam USVI etc), India, Vietnam, China, Hong Kong, Cambodia, France, Laos, Haiti, South Korea, Iran, Greece, Italy, Poland, Thailand, Yugoslavia, Holand-Netherlands, Ireland, Scotland, Hungary, Panama.

country of birth mother: India, Mexico, United-States, Puerto-Rico, Dominican-Republic, England, Honduras, Peru, Guatemala, Columbia, El-Salvador, Philippines, France, Ecuador, Nicaragua, Cuba, Outlying-U S (Guam USVI etc), Jamaica, South Korea, China, Germany, Yugoslavia, Canada, Vietnam, Japan, Cambodia, Ireland, Laos, Haiti, Portugal, Taiwan, Holand-Netherlands, Greece, Italy, Poland, Thailand, Trinidad&Tobago, Hungary, Panama, Hong Kong, Scotland, Iran.

country of birth self: United-States, Mexico, Puerto-Rico, Peru, Canada, South Korea, India, Japan, Haiti, El-Salvador, Dominican-Republic, Portugal, Columbia, England, Thailand, Cuba, Laos, Panama, China, Germany, Vietnam, Italy, Honduras, Outlying-U S (Guam USVI etc), Hungary, Philippines, Poland, Ecuador, Iran, Guatemala, Holand-Netherlands, Taiwan, Nicaragua, France, Jamaica, Scotland, Yugoslavia, Hong Kong, Trinidad&Tobago, Greece, Cambodia, Ireland.

citizenship: Native- Born in the United States, Foreign born- Not a citizen of U S , Native-Born in Puerto Rico or U S Outlying, Native- Born abroad of American Parent(s), Foreign born- U S citizen by naturalization.

own business or self employed: 0, 2, 1.

fill inc questionnaire for veteran's admin: Not in universe, Yes, No.

veterans benefits: 0, 2, 1.

weeks worked in year: continuous.

year: 94, 95.