Low Level Design (LLD)

Adult Census Income Prediction

Revision Number: 0

Manojprabakaran.M

Document Version Control

Date Issued	Version	Description	Author
12th Dec 2023	1	submission	Manojprabakaran

Contents

Abstract

Our Adult census income prediction project leverages machine learning to classify individuals as having an income greater or less than 50k. Using a diverse range of features, including demographics, education, and work-related data, our model is designed for international applicability. We explore various classification algorithms and employ ethical data practices to ensure fairness. Our project offers insights into socioeconomic disparities and provides a valuable tool for informed decision-making and resource allocation worldwide.

1 Introduction

1.1 Why this Low-Level Design Document?

The purpose of this document is to present a detailed description of the Adult Census Income Prediction. It will explain the purpose and features of the system, the interfaces of the system, what the system will do, the constraints under which it must operate and how the system will react to external stimuli.

The main objective of the project is to predict if person income is greater than or equal to 50k or less than 50k . Dataset consist of all the valuable information.

An Adult census income prediction contains a following features, such as:

- Age ● Workclass.
- Education.
- Occupation.
- Sex
- Race.
- Capital gain
- Capital loss
- Hour per week
- Country

1.2 Scope

This software system will be a Web application This system will be designed to detect the Income of the user based on various senerio, Whether person income is below or equal 50k or more than 50k. This system is designed to predict the income for many prospects for Risk management, For businesses purpose , investment decisions etc .

1.3 Out of Scope

Delineate specific activities, capabilities, and items that are out of scope for the project. Many things are not under the scope such as cities , married status etc

2 Technical specifications

2.1 Dataset

Column ID	Column Name	Data type	Values type	Description
0	age	int64	Continous	Age of person
1	workclass	object	Discrete	Workclass of person
2	fnlwgt	int64	Continous	Final weight
3	education	object	Discrete	Education Degree of person
4	education.num	int64	Continous	Number of years of education
5	marital.status	object	Discrete	Marital status of person
6	occupation	object	Discrete	Occupation of person
7	relationship	object	Discrete	Relationship of person
8	race	object	Discrete	Race of person
9	sex	object	Discrete	Sex of person
10	capital.gain	int64	Continous	Capital gain of person
11	capital.loss	int64	Continous	Capital loss of person
12	hours.per.week	int64	Continous	Number of hours per week
13	native.country	object	Discrete	Native country of person
14	income	object	Discrete	Income category of person

2.1.1 Adult census income prediction dataset overview

Consists of various attributes age workclaass, race, sex, occupation ,work hour per week and country. Hence we are performing the classification problem and our target column is salary which is a binary class ,

There are a total of 22792 rows in the training set and 9769 rows in the test set.and rest of columns are remain same

- Training data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	age	workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race	sex	capital-gain	capital-loss	hours-per-we	country	salary
2	34	Self-emp-not-inc	56460	HS-grad	9	Married-civ-spouse	Farming-fishing	Wife	White	Female	0	2179	12	United-States	<=50K
3	48	Self-emp-not-inc	243631	Some-college	10	Married-civ-spouse	Craft-repair	Husband	Amer-Indian-Es	Male	7688	0	40	United-States	>50K
4	23	State-gov	56402	Some-college	10	Married-civ-spouse	Prof-specialty	Husband	White	Male	0	0	30	United-States	<=50K
5	56	Local-gov	255406	HS-grad	9	Divorced	Exec-manageria	Not-in-family	White	Female	0	0	40	United-States	<=50K
6	17	Private	297246	11th	7	Never-married	Priv-house-serv	Own-child	White	Female	0	0	9	United-States	<=50K
7	34	Private	152453	12th	8	Married-civ-spouse	Craft-repair	Husband	White	Male	0	0	40	Guatemala	<=50K
8	37	Private	140673	Some-college	10	Married-civ-spouse	Sales	Wife	White	Female	0	0	40	United-States	>50K
9	49	Self-emp-not-inc	162856	Some-college	10	Divorced	Exec-manageria	Not-in-family	Amer-Indian-Es	Female	0	0	40	United-States	<=50K
10	55	Private	132887	HS-grad	9	Married-civ-spouse	Transport-movin	Husband	Black	Male	3411	0	40	Jamaica	<=50K
11	25	Private	131463	Bachelors	13	Married-civ-spouse	Prof-specialty	Husband	White	Male	0	0	40	United-States	<=50K
12	38	Private	255621	HS-grad	9	Married-civ-spouse	Other-service	Husband	White	Male	0	0	40	United-States	<=50K
13	37	Private	38468	Some-college	10	Never-married	Adm-clerical	Not-in-family	White	Female	0	0	40	United-States	<=50K
14	45	Private	88500	Some-college	10	Married-civ-spouse	Machine-op-insp	Husband	White	Male	0	0	44	United-States	>50K
15	51	Local-gov	26832	HS-grad	9	Married-civ-spouse	Craft-repair	Husband	White	Male	0	0	40	United-States	>50K
16	47	Private	52795	Some-college	10	Married-civ-spouse	Machine-op-insp	Husband	White	Male	0	0	46	United-States	<=50K
17	34	Private	314646	9th	5	Married-civ-spouse	Craft-repair	Husband	White	Male	0	0	40	United-States	<=50K
18	27	Local-gov	322208	Some-college	10	Never-married	Transport-movin	Own-child	White	Male	0	0	40	United-States	<=50K
19	45	Private	140644	HS-grad	9	Married-civ-spouse	Machine-op-insp	Husband	White	Male	0	0	40	United-States	<=50K
20	21	Private	73679	Some-college	10	Never-married	Exec-manageria	Own-child	White	Female	0	0	40	United-States	<=50K
21	23	Private	85088	Some-college	10	Never-married	Other-service	Own-child	White	Female	0	0	37	United-States	<=50K
22	46	Private	133616	Some-college	10	Divorced	Adm-clerical	Unmarried	White	Female	0	0	40	United-States	<=50K
23	22	Private	192455	HS-grad	9	Never-married	Adm-clerical	Unmarried	White	Female	0	0	40	United-States	<=50K
24	31	Private	173350	HS-grad	9	Never-married	Other-service	Not-in-family	White	Female	0	0	30	United-States	<=50K
25	33	Private	199046	Bachelors	13	Never-married	Sales	Not-in-family	White	Female	0	0	40	United-States	<=50K
26	55	Self-emp-not-inc	322691	Masters	14	Married-civ-spouse	Sales	Husband	White	Male	3103	0	55	United-States	>50K
27	24	Private	278130	Assoc-acdm	12	Never-married	Adm-clerical	Not-in-family	White	Male	0	0	40	United-States	<=50K
28	33	Private	127215	Some-college	10	Married-civ-spouse	Sales	Husband	White	Male	0	0	48	United-States	>50K
29	43	Private	269015	HS-grad	9	Married-civ-spouse	Machine-op-insp	Husband	Black	Male	0	0	40	Germany	>50K
30	46	Private	251786	1st-4th	2	Separated	Other-service	Not-in-family	White	Female	0	0	40	Mexico	<=50K
31	18	Private	115839	12th	8	Never-married	Adm-clerical	Not-in-family	White	Female	0	0	30	United-States	<=50K
32	23	Private	277328	Some-college	10	Never-married	Handlers-cleane	Own-child	White	Male	0	0	32	Cuba	<=50K
33	44	Federal-gov	244054	Some-college	10	Divorced	Exec-manageria	Unmarried	White	Male	0	0	60	United-States	>50K
34	30	Self-emp-not-inc	173854	HS-grad	9	Married-civ-spouse	Craft-repair	Husband	White	Male	0	0	60	United-States	<=50K
35	22	?	376277	Some-college	10	Divorced	?	Not-in-family	White	Female	0	0	35	United-States	<=50K

training

- Test data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	age	workclass	fnlwgt	education	education-	marital-sta	occupatio	relationshi	race	sex	capital-gai	capital-los	hours-per-	country	salary
2	27	Private	160178	Some-coll	10	Divorced	Adm-cleri	Not-in-far	White	Female	0	0	38	United-Sta	<=50K
3	45	State-gov	50567	HS-grad	9	Married-c	Exec-man	Wife	White	Female	0	0	40	United-Sta	<=50K
4	29	Private	185908	Bachelors	13	Married-c	Exec-man	Husband	Black	Male	0	0	55	United-Sta	>50K
5	30	Private	190040	Bachelors	13	Never-ma	Machine-c	Not-in-far	White	Female	0	0	40	United-Sta	<=50K
6	29	Self-emp-	189346	Some-coll	10	Divorced	Craft-repa	Not-in-far	White	Male	2202	0	50	United-Sta	<=50K
7	51	Private	108435	Masters	14	Married-c	Prof-speci	Husband	White	Male	0	0	47	United-Sta	>50K
8	58	Self-emp-	93664	HS-grad	9	Married-c	Exec-man	Husband	White	Male	15024	0	60	United-Sta	>50K
9	22	Private	148431	HS-grad	9	Never-ma	Adm-cleri	Not-in-far	Other	Female	0	0	40	United-Sta	<=50K
10	50	Private	313321	Assoc-acd	12	Divorced	Sales	Not-in-far	White	Female	0	0	40	United-Sta	<=50K
11	50	Private	71417	HS-grad	9	Married-c	Craft-repa	Husband	White	Male	3103	0	40	United-Sta	>50K
12	43	Self-emp-	241895	Bachelors	13	Never-ma	Sales	Not-in-far	White	Male	0	0	42	United-Sta	<=50K
13	37	Private	271013	HS-grad	9	Divorced	Handlers-c	Not-in-far	White	Male	0	0	40	United-Sta	<=50K
14	41	Private	152529	Bachelors	13	Never-ma	Prof-speci	Not-in-far	White	Male	0	0	40	United-Sta	>50K
15	46	Private	116952	7th-8th	4	Married-c	Prof-speci	Wife	White	Female	0	0	45	United-Sta	<=50K
16	33	Private	226629	Some-coll	10	Married-c	Other-ser	Wife	White	Female	0	0	35	United-Sta	<=50K
17	45	Private	161819	11th	7	Separated	Adm-cleri	Unmarrie	Black	Female	0	0	25	United-Sta	<=50K
18	43	Private	99212	Assoc-voc	11	Married-c	Exec-man	Husband	White	Male	3103	0	48	United-Sta	>50K
19	52	Self-emp-	179951	Prof-scho	15	Married-c	Exec-man	Husband	White	Male	0	0	40	United-Sta	<=50K
20	38	Federal-g	115433	Assoc-acd	12	Married-c	Adm-cleri	Wife	White	Female	7688	0	33	United-Sta	>50K
21	32	Self-emp-	267161	Bachelors	13	Divorced	Exec-man	Unmarrie	Black	Female	0	0	30	United-Sta	<=50K
22	40	Private	439919	9th	5	Married-c	Machine-c	Husband	White	Male	0	0	40	Mexico	<=50K
23	31	State-gov	93589	HS-grad	9	Divorced	Protective	Own-child	Other	Male	0	0	40	United-Sta	<=50K
24	51	Private	95329	Masters	14	Divorced	Protective	Unmarrie	White	Male	0	0	40	United-Sta	<=50K
25	23	Private	227594	HS-grad	9	Never-ma	Adm-cleri	Not-in-far	White	Female	0	0	30	United-Sta	<=50K
26	29	Private	50295	Some-coll	10	Married-c	Exec-man	Husband	White	Male	0	0	48	United-Sta	<=50K
27	32	Local-gov	159187	Bachelors	13	Never-ma	Prof-speci	Not-in-far	White	Male	0	0	35	United-Sta	<=50K
28	59	Private	167963	HS-grad	9	Never-ma	Adm-cleri	Unmarrie	White	Male	0	0	40	United-Sta	<=50K
29	32	Private	237903	Bachelors	13	Never-ma	Prof-speci	Unmarrie	White	Female	0	0	40	United-Sta	<=50K

test

2.1.2 Input schema

Feature name	Datatype	Size	Null/Required
Age	int	NIL	Required
Work class	str	NIL	Required
Education Num	int	NIL	Required
Occupation	str	NIL	Required
Race	str	NIL	Required
Sex	str	NIL	Required
Capital Gain	Int	NIL	Required
Capital loss	Int	NIL	Required
Hour per week	Int	NIL	Required
Country	str	NIL	Required

2.2 Predicting Income

- The system displays the input menu for Age of the user .
- The system displays the dropdown menu to select any choice of them from workclass.
- The system presents the set of inputs required from the user.
- The User chooses the occupation by clicking one of the availabilies.
- The User heave to selects from race column by given information.
- The system presents the set of inputs required from the user.
- The user gives required information.
- The system should be able to predict whether user Income is below or equal to 50k or above 50k

2.3 Logging

We should be able to log every activity done by the user.

- The System identifies at what step logging required
- The System should be able to log each and every system flow.
- Developers can choose logging methods. You can choose database logging/ File logging as well.

2.4 Database

System needs to store every request into the database and we need to store it in such a way that it is easy to retrain the model as well.

1. The User should gives all the required information.
2. Directly connected with the database MongoDB
3. The system stores each and every data given by the user or received on request to the database.

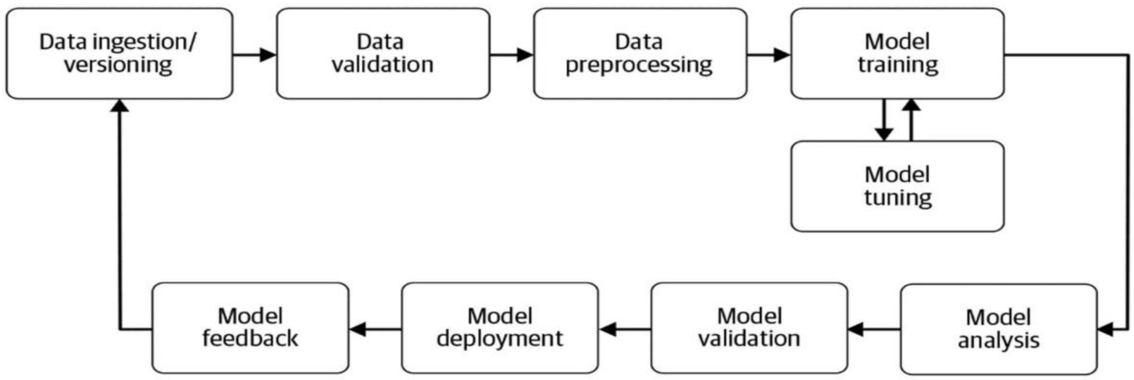
2.5 Deployment

1. AWS

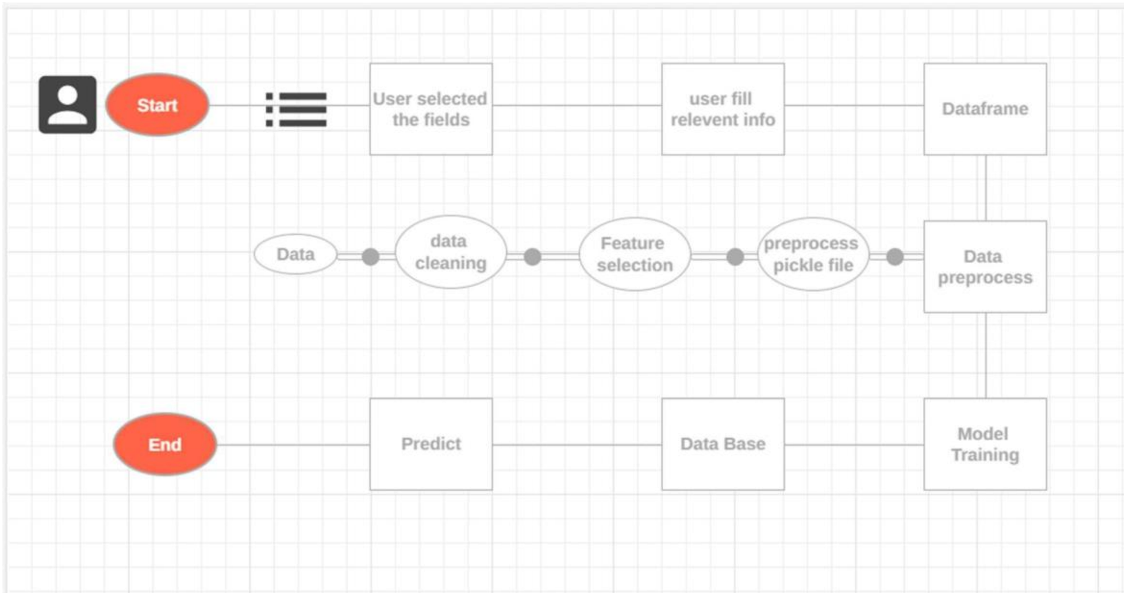
3 Technology stack

Front End	HTML/CSS
Backend	Python /Flask
Database	MongoDB
Deployment	AWS

4 Model training/validation workflow



5 User I/O workflow



6 Test cases

Test case	Steps to perform test case	Module	Pass/Fail