TELECOM CUSTOMER CHURN PREDICTION

Presented by: E. Alekseeva, S. Kazakova, G. Kudabaeva,
E. Nezhurina, A. Unacheva, E. Zolotukhina

TABLE OF CONTENTS

Why it is important to analyze customer churn?

- Improves Customer Retention Strategies.
- Strengthens Competitive Advantage
- Enhances Revenue Stability
- Provides Early Warning for Business Health
- Supports Resource Allocation Efficiency
- Informs Product and Service Improvement

Summary of previous results

	Logistic Regression	Random Forest	SVM	Gradient Boosting
Accuracy score	78.7%	76.6%	78.8%	78.2%
Precision score	63%	61%	68%	69%
Recall score	44%	45%	36%	43%
F1-score	52%	52%	47%	53%

Source: Customer Churn Prediction (AIIoT 2024)

About the data

The dataset contains 243,553 rows of customer data from four major telecom partners of India: Airtel, Reliance Jio, Vodafone, and BSNL.

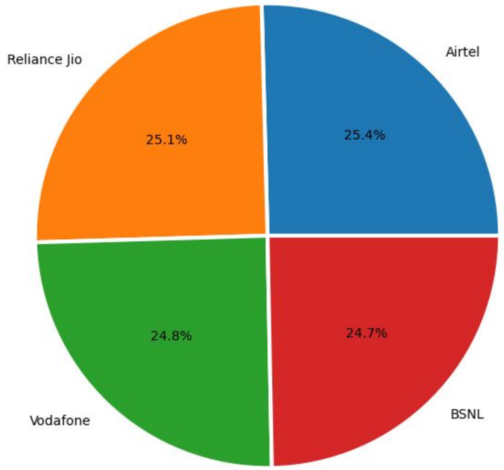


Fig.1. Pie chart showing the share of companies in the sample.

Variables

The dataset includes various demographic, location, and usage pattern variables for each customer.

Target variable

“Churn” - binary variable indicating whether the customer has churned or not.

0	194726	79%
1	48827	21%

Descriptive statistics

	customer_id	age	pincode	num_dependents	estimated_salary	calls_made	sms_sent	data_used	churn
count	243553.000000	243553.000000	243553.000000	243553.000000	243553.000000	243553.000000	243553.000000	243553.000000	243553.000000
mean	121777.000000	46.077609	549501.270541	1.997500	85021.137839	49.010548	23.945404	4993.186025	0.200478
std	70307.839393	16.444029	259808.860574	1.414941	37508.963233	29.453556	14.733575	2942.019547	0.400359
min	1.000000	18.000000	100006.000000	0.000000	20000.000000	-10.000000	-5.000000	-987.000000	0.000000
25%	60889.000000	32.000000	324586.000000	1.000000	52585.000000	24.000000	11.000000	2490.000000	0.000000
50%	121777.000000	46.000000	548112.000000	2.000000	84990.000000	49.000000	24.000000	4987.000000	0.000000
75%	182665.000000	60.000000	774994.000000	3.000000	117488.000000	74.000000	36.000000	7493.000000	0.000000
max	243553.000000	74.000000	999987.000000	4.000000	149999.000000	108.000000	53.000000	10991.000000	1.000000

Original data

- **14 variables**
  - 4 categorical
  - 9 numerical
  - 1 binary
- **243553 observations**
- **No missing values**

Processed data

- Negative values were **replaced with zeroes**
- The **estimated_salary** was **converted to logarithm**
- **Categorical variables were encoded** using pd.get_dummies()
- **SMOTE** method for the training sample
- Data normalization - **Standard Scaler**

Graphic description

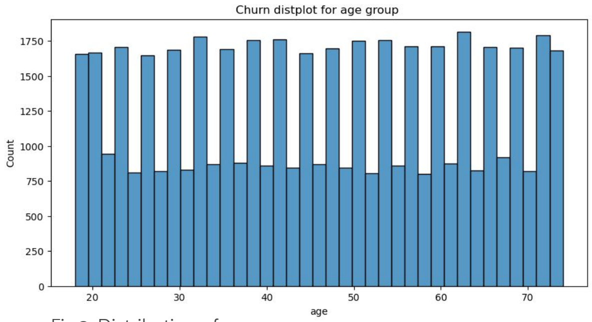


Fig.2. Distribution of age.

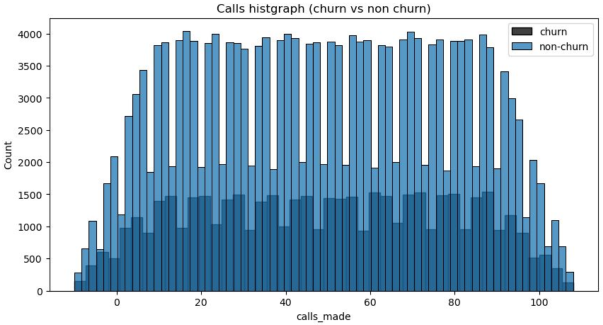


Fig.3. Distribution of calls of churn and non-churn clients.

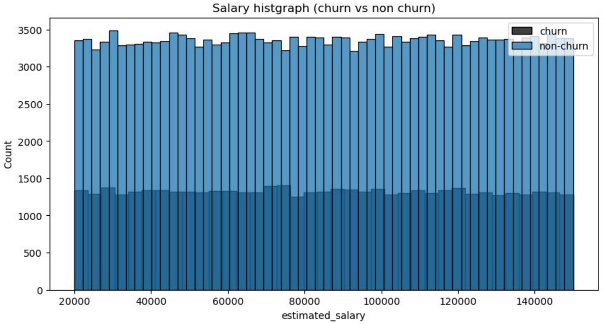


Fig.4. Distribution of salary of churn and non-churn clients.

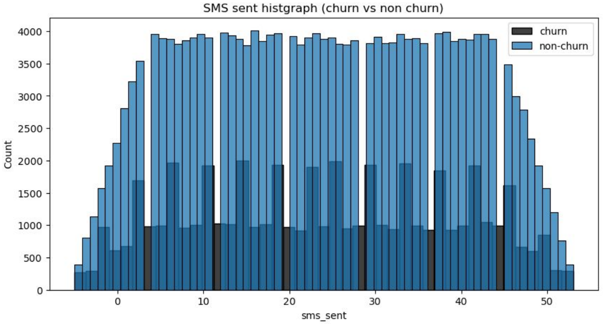


Fig.5. Distribution of SMS sent of churn and non-churn clients.

Graphic description

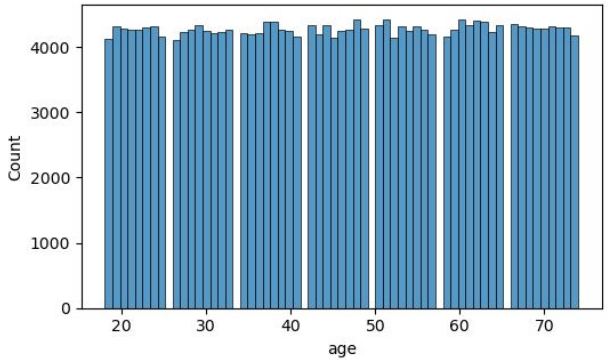


Fig.6. Distribution of age.

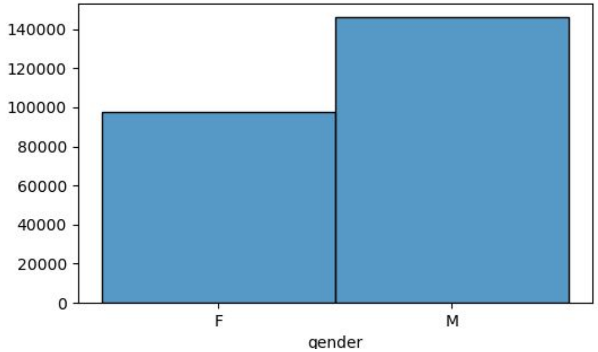


Fig.7. Histogram of gender.

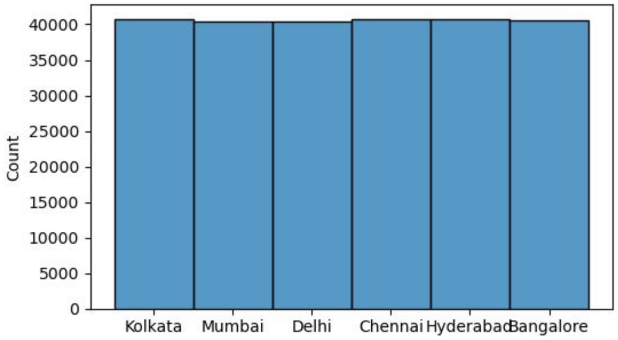


Fig.8. Histogram of cities.

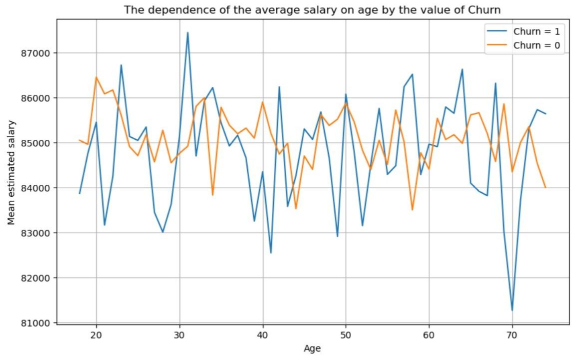


Fig.9. Dependence of the average salary on age by the value of Churn.

Logit

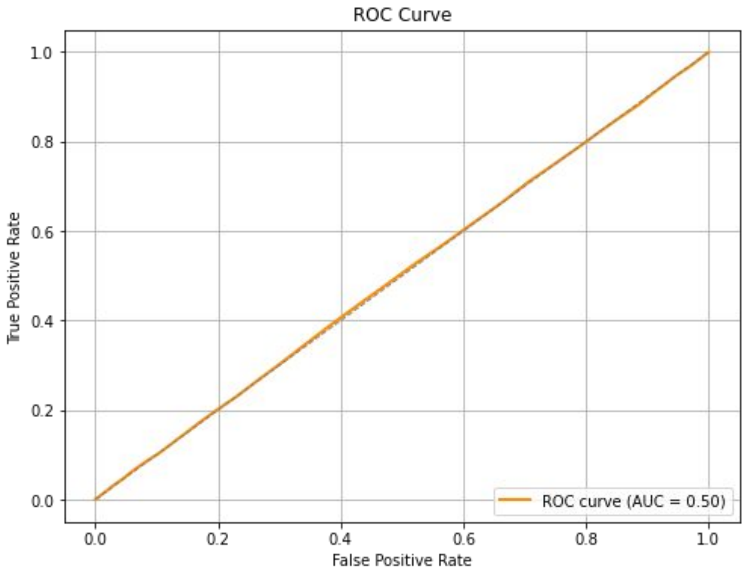


Fig.10. ROC-curve for logit model.

	Logit without tuning	Logit with ADASYN
Accuracy score	80%	51%
Precision score	0%	20%
Recall score	0%	49%
F1-score	0%	29%

Logit: marginal effects

Variable	Coefficient	Marginal effects	P-value
Age	-4.19e-06	1.88e-05	0.804
Gender	**2.35e-06**	**0.0429**	**0.000**
Log Estimated Salary	4.57e-11	0.0003	0.944
Vodafone Company	**1.86e-06**	**0.1290**	**0.000**
Calls made	8.62e-08	0.0003	0.889
Sms sent	-1.16e-07	-0.0031	0.104
Registration year	**-6.79e-07**	**-0.0186**	**0.000**
Number of dependents	**-4.17e-06**	**-0.0100**	**0.000**

Random Forest

	Random Forest	Random Forest with SMOTE
Accuracy score	80%	78%
Precision score	50%	20%
Recall score	0%	19%
F1-score	0%	3.5%

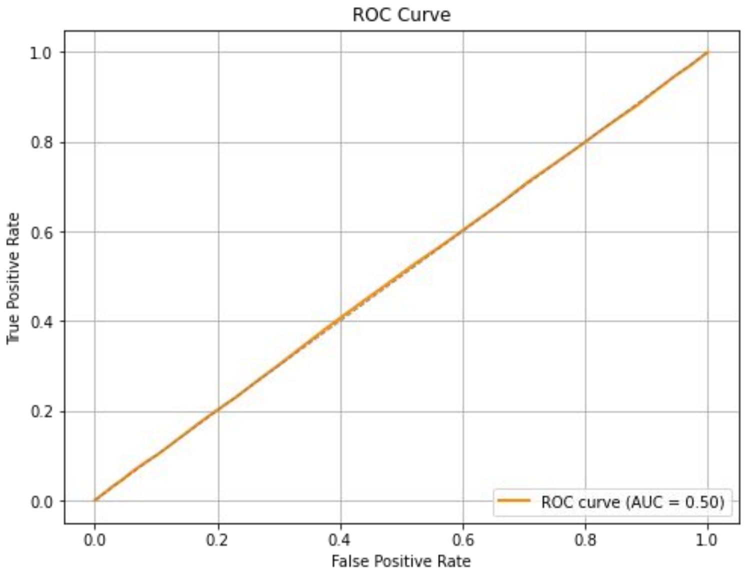


Fig.11. ROC curve for Random Forest model.

Gradient Boosting

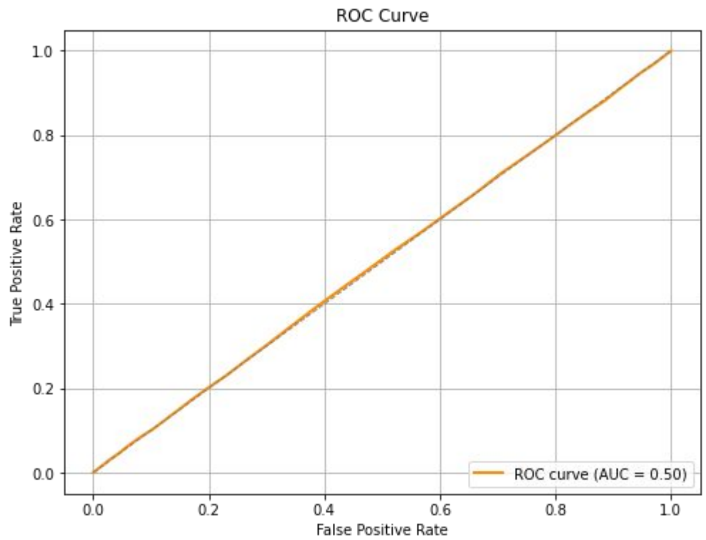


Fig.12. ROC curve for GB model.

	GB without tuning	GB with tuned parameters
Accuracy score	80%	79%
Precision score	0%	19%
Recall score	0%	0.8%
F1-score	0%	1.5%

XGBoost

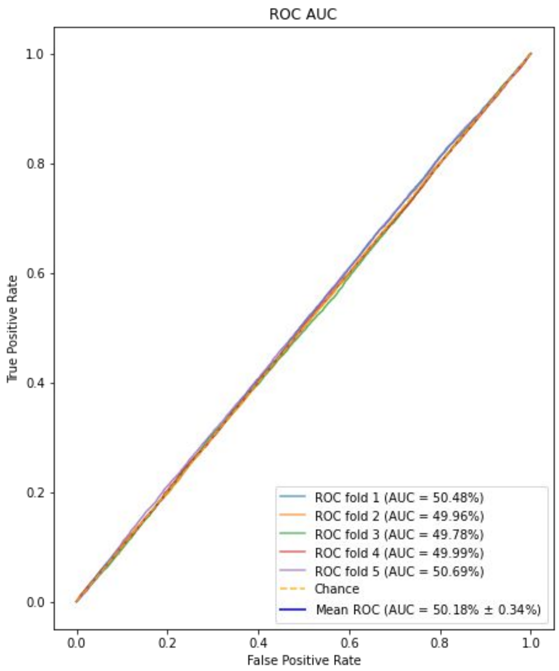


Fig.13. ROC curve for XGboost model.

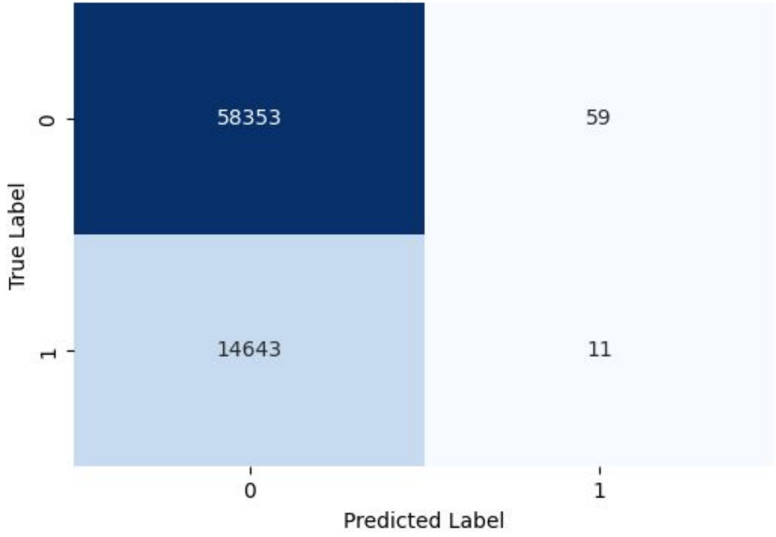
	XgBoost
Accuracy score	80%
Precision score	21%
Recall score	0.07%
F1-score	0.14%

Stacking

	Base models: RF and GB Classifier	Base models: RF and GB Classifier with SMOTE	Base models: RF and XGBoost	Base models: RF and XGBoost with ADASYN
Accuracy score	80%	71%	79%	80%
Precision score	16%	19%	12%	16%
Recall score	0.07%	14%	0.08%	0.08%
F1-score	0.1%	16%	0.16%	0.14%

Stacking

Fig.14. Confusion Matrix for Stacking Model with ADASYN & XGBoost



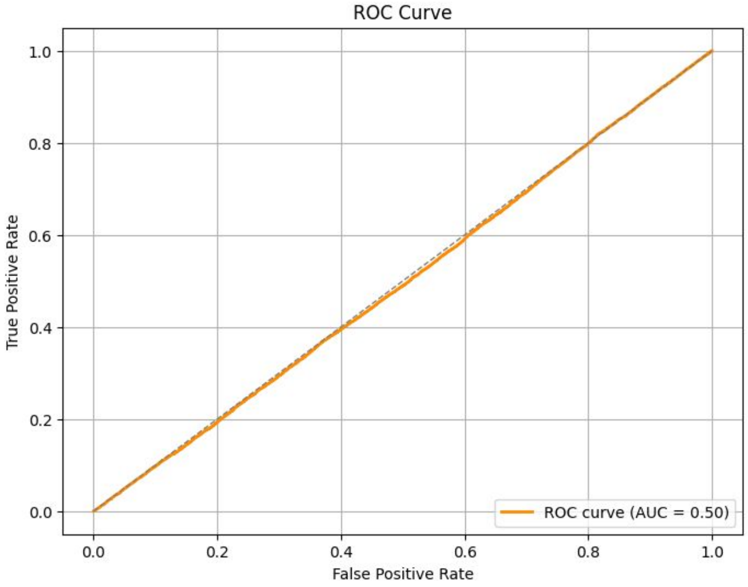


Fig.15. ROC curve for Stacking.

Summary

- **Age** and **estimated salary** of customers **do not affect customer churn**
- The amount of **calls made** and **sms sent do not affect customer churn**
- The **later the number was registered**, the **lower** the **probability** of customer **churn**
- The **higher the number of dependents**, the **lower** the **probability** of customer **churn**

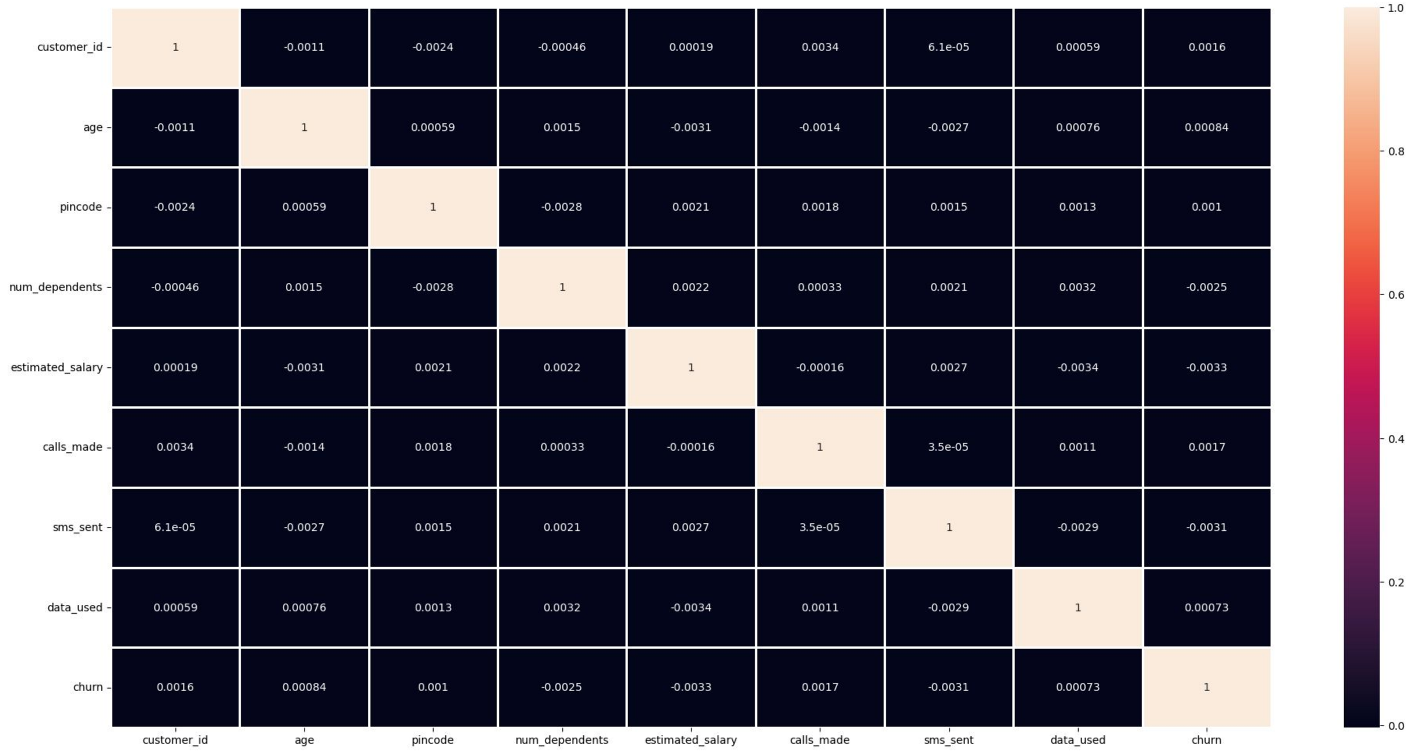
Best models: Gradient Boosting tuned and Stacking with RF and XGB classifier

	Logit with ADASYN	Random Forest tuned	Gradient Boosting tuned	XGBoost	Stacking with RF & XGB ADASYN
Accuracy score	80%	78%	79%	79%	80%

THANK YOU FOR ATTENTION!

Appendix

Appendix 1. Correlation matrix



Appendix

Appendix 2. Table of marginal effects

Dep. Variable: churn
Method: dydx
At: mean

	dy/dx	std err	z	P>|z|	[0.025	0.975]
age	1.881e-05	7.58e-05	0.248	0.804	-0.000	0.000
pincode	9.595e-09	4.71e-09	2.037	0.042	3.64e-10	1.88e-08
num_dependents	-0.0100	0.001	-11.315	0.000	-0.012	-0.008
estimated_salary	-3.604e-08	5.32e-08	-0.677	0.498	-1.4e-07	6.82e-08
calls_made	0.0003	0.002	0.140	0.889	-0.004	0.004
sms_sent	-0.0031	0.002	-1.626	0.104	-0.007	0.001
data_used	0.0016	0.002	0.806	0.420	-0.002	0.005
registration_year	-0.0186	0.001	-14.329	0.000	-0.021	-0.016
registration_month	-0.0026	0.000	-7.147	0.000	-0.003	-0.002
telecom_partner_BSNL	0.1319	0.003	38.892	0.000	0.125	0.139
telecom_partner_Reliance Jio	0.1338	0.003	39.499	0.000	0.127	0.140
telecom_partner_Vodafone	0.1290	0.003	37.992	0.000	0.122	0.136
gender_M	0.0429	0.003	16.684	0.000	0.038	0.048
state_Arunachal Pradesh	1.0030	0.008	132.937	0.000	0.988	1.018
state_Assam	1.0118	0.008	133.842	0.000	0.997	1.027
state_Bihar	1.0020	0.008	131.438	0.000	0.987	1.017
state_Chhattisgarh	0.9927	0.008	130.169	0.000	0.978	1.008
state_Goa	1.0033	0.008	132.206	0.000	0.988	1.018
state_Gujarat	1.0009	0.008	130.627	0.000	0.986	1.016
state_Haryana	1.0094	0.008	134.393	0.000	0.995	1.024
state_Himachal Pradesh	1.0136	0.007	135.235	0.000	0.999	1.028
state_Jharkhand	1.0172	0.007	136.419	0.000	1.003	1.032
state_Karnataka	1.0141	0.007	135.592	0.000	0.999	1.029
state_Kerala	1.0032	0.008	132.206	0.000	0.988	1.018
state_Madhya Pradesh	1.0099	0.008	134.506	0.000	0.995	1.025
state_Maharashtra	1.0109	0.008	134.767	0.000	0.996	1.026
state_Manipur	1.0094	0.008	133.681	0.000	0.995	1.024
state_Meghalaya	1.0021	0.008	131.893	0.000	0.987	1.017
state_Mizoram	1.0189	0.007	136.045	0.000	1.004	1.034
state_Nagaland	1.0053	0.008	132.932	0.000	0.991	1.020
state_Odisha	1.0083	0.008	133.453	0.000	0.993	1.023
state_Punjab	0.9939	0.008	129.623	0.000	0.979	1.009
state_Rajasthan	1.0058	0.008	133.013	0.000	0.991	1.021
state_Sikkim	1.0006	0.008	131.687	0.000	0.986	1.015
state_Tamil Nadu	1.0058	0.008	132.987	0.000	0.991	1.021
state_Telangana	1.0055	0.008	132.707	0.000	0.991	1.020
state_Tripura	1.0072	0.008	133.359	0.000	0.992	1.022
state_Uttar Pradesh	1.0092	0.008	134.214	0.000	0.994	1.024
state_Uttarakhand	1.0151	0.007	135.953	0.000	1.000	1.030
state_West Bengal	0.9903	0.008	128.426	0.000	0.975	1.005
city_Chennai	0.2213	0.004	53.359	0.000	0.213	0.229
city_Delhi	0.2224	0.004	53.449	0.000	0.214	0.231
city_Hyderabad	0.2284	0.004	54.982	0.000	0.220	0.236
city_Kolkata	0.2249	0.004	54.211	0.000	0.217	0.233
city_Mumbai	0.2212	0.004	53.197	0.000	0.213	0.229
log_estimated_salary	0.0003	0.004	0.071	0.944	-0.008	0.008