AgroFides

Predictive Modeling Field Project
Brandeis International Business School

Xinyu Cao, MA-DA
Silin Chen, MSF
Guangran Zhu, MSF

Background & Objectives

Business challenges:

- Find the best model that can predict whether an investment on specific farmers will return profits.

Goal of project:

- Assess the completeness of the dataset
- Identify key factors that predict the creditworthiness of farmers in Ghana
- Build a model that predicts the credit score or other forms of creditworthiness of the farmers.
- Test the model, reporting out on model diagnostics

Our Process

Step 1 Building a Theoretical Model

Step 2 Data Audit

Step 3 Data Cleaning

Step 4 Train and Test Model

Step 5 Model Comparison

Step 1:
Building A Theoretical Model

What We Ultimately Want To Do

X → Y

FARM OR FARMER CHARACTERISTICS | PREDICTS | FARMER PROFITABILITY

We researched what type of metrics were best to measure profitability

X → Y

FARM OR FARMER CHARACTERISTICS | PREDICTS | FARMER PROFITABILITY

References

Zhu X, Li ZN. (2007): Farmer credit is necessary for access to working capital and credit loans offered by financial institutions.

Wang TC, Chen YH (2006): Small and medium sized customer credit ratings may also be evaluated using the "5C principle": **Character**, **Capital**, **Capacity**, Collateral and Condition of Business.

Klaus Maurer (2014): The risks in agricultural finance comprise to a considerable extent common risks associated with the **viability** of the farm business and the farmer's character, not much different from the risks of micro and small businesses in other economic sectors .

OECD (2009): **five major sources of risk** in agriculture can be defined: production risk, market risk, financial risk, legal and environment risk, human resource risks

We researched what type of characteristics would be the best predictors

X	→	Y
FARM OR FARMER CHARACTERISTICS	PREDICTS	FARMER PROFITABILITY

References

Bojnec and Latruffe (2013) find that **small farms** are less technically efficient.

Kauffman and Tauerand Haden and Johnson (1985) used **expenditures** on hired labor as an explanatory variable.

Johnson, Prescott, Banker, and Morehart; Reimund and Somwaru; and Strickland: **characteristics** such as farm size, location, and **cash grain production** were positively related to a measure of profit. Conversely, livestock production and age of operator were negatively related to a measure of farm profit.

Reinsel and Joseph found that commodities produced, location, size of operation, management, and **natural phenomena** are factors that cause returns to vary.

Our Hypothetical Model

X → Y

Constructs From Research:

Info (Farm & Farmer) + Operation (Human & Machine) + Environment (Weather & Economics) + Marketing = Farmer Probability

Step 2:
Data Audit

What Is A Data Audit?

What We Do:

The Process

Use variables that match our theoretical model

Use variables to construct a new feature that matches our theoretical model

What We Get Out Of It:

The Deliverable

Taking your data and fitting it to our theoretical model

Our Hypothetical Model With Variables From Your Data Set

Constructs From Research:

Info (Farm & Farmer) + Operation (Human & Machine) + Environment (Weather & Economics) + Marketing = Farmer Profitability

Proxy From Data Sets:

E.g. Farmers education level, Size of farm + E.g. Irrigate, labor_days per week + E.g. dry spell or drought in last two years? + E.g. Distance to market in mile = See Next Slide

Data We Still need:

Operation (Mach) E.g. Any automated machinery? + Environment (Econ) E.g. Government subsidy? + Marketing E.g. Storage & Transportation? = See Next Slide

How We Calculated Our Target Variable

Creditworthiness

(e.g., Sales / Operation Cost)

target_2019 = (sales in 2019 / operation_cost in 2019)

target_2018 = (sales in 2018 / operation_cost in 2018)

Step 3:
Data Cleaning

Deliverable 2 - Data Cleaning

Step 1	Convert all data to float
Step 2	Fill NA using mean/mode or percentage
Step 3	Add target variable (Profitability)
Step 4	Get dummy variables
Step 5	Rescaling
Step 6	Feature engineering

Fill all NA Data

Reason:

- Cannot run models when data missing

Method:

- Using mean, mode or based on percentage

Image Of Data Pre filling NA

	age	gender	household	marriage	wives	children	childEdu	farmerEdu	farmPrimary	farmNum	farmType	farmSize
0	4.007333	M	1	1.0	2.0	11.0	NaN	0.0	1.0	2.0	1.0	4.0
1	3.713572	M	1	1.0	2.0	7.0	NaN	1.0	1.0	1.0	1.0	4.0
2	4.060443	M	1	1.0	2.0	9.0	NaN	0.0	1.0	3.0	1.0	4.0
3	3.806663	F	1	1.0	0.0	8.0	NaN	0.0	1.0	2.0	1.0	3.0
4	4.007333	F	1	1.0	0.0	9.0	NaN	0.0	1.0	1.0	1.0	1.0
...	...	...	...	...	...	...	...	...	...	...	...	...
507	4.248495	M	0	1.0	1.0	6.0	NaN	2.0	1.0	2.0	2.0	2.0
508	4.043051	F	1	1.0	1.0	4.0	NaN	2.0	1.0	1.0	2.0	14.0
509	4.110874	M	0	1.0	1.0	4.0	NaN	2.0	1.0	1.0	2.0	3.0
510	4.025352	F	1	1.0	1.0	7.0	NaN	0.0	1.0	1.0	2.0	6.0
511	3.555348	M	1	1.0	1.0	4.0	4.0	2.0	1.0	2.0	2.0	6.0

512 rows × 44 columns

Image Of Data POST filling NA

	age	gender	household	marriage	wives	children	childEdu	farmerEdu	farmPrimary	farmNum	farmType	farmSize
0	4.007333	M	1	1.0	2.0	11.0	0.0	0.0	1.0	2.0	1.0	4.0
1	3.713572	M	1	1.0	2.0	7.0	0.0	1.0	1.0	1.0	1.0	4.0
2	4.060443	M	1	1.0	2.0	9.0	0.0	0.0	1.0	3.0	1.0	4.0
3	3.806663	F	1	1.0	0.0	8.0	0.0	0.0	1.0	2.0	1.0	3.0
4	4.007333	F	1	1.0	0.0	9.0	0.0	0.0	1.0	1.0	1.0	1.0
...	...	...	...	...	...	...	...	...	...	...	...	...
507	4.248495	M	0	1.0	1.0	6.0	0.0	2.0	1.0	2.0	2.0	2.0
508	4.043051	F	1	1.0	1.0	4.0	0.0	2.0	1.0	1.0	2.0	14.0
509	4.110874	M	0	1.0	1.0	4.0	0.0	2.0	1.0	1.0	2.0	3.0
510	4.025352	F	1	1.0	1.0	7.0	1.0	0.0	1.0	1.0	2.0	6.0
511	3.555348	M	1	1.0	1.0	4.0	4.0	2.0	1.0	2.0	2.0	6.0

512 rows × 44 columns

Create Dummy Variables

Reason:

- Transfer all data to numerical

Method:

- get_dummies()

Image Of Data Pre dummy coding

	gender	household	marriage	farmerEdu	farmPrimary	farmType	otherIncome	drought	fires	flood	wind
0	M	1	1.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0
1	M	1	1.0	1.0	1.0	1.0	1.0	0.0	0.0	1.0	0.0
2	M	1	1.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0
3	F	1	1.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0
4	F	1	1.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...
507	M	0	1.0	2.0	1.0	2.0	1.0	0.0	0.0	0.0	1.0
508	F	1	1.0	2.0	1.0	2.0	0.0	1.0	0.0	0.0	0.0
509	M	0	1.0	2.0	1.0	2.0	0.0	0.0	0.0	0.0	0.0
510	F	1	1.0	0.0	1.0	2.0	0.0	1.0	0.0	0.0	0.0
511	M	1	1.0	2.0	1.0	2.0	0.0	0.0	1.0	1.0	0.0

512 rows × 23 columns

Image Of Data POST dummy coding

	household	marriage	farmerEdu	farmPrimary	farmType	otherIncome	drought	fires	flood	wind
0	1	1.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0
1	1	1.0	1.0	1.0	1.0	1.0	0.0	0.0	1.0	0.0
2	1	1.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0
3	1	1.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0
4	1	1.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...
507	0	1.0	2.0	1.0	2.0	1.0	0.0	0.0	0.0	1.0
508	1	1.0	2.0	1.0	2.0	0.0	1.0	0.0	0.0	0.0
509	0	1.0	2.0	1.0	2.0	0.0	0.0	0.0	0.0	0.0
510	1	1.0	0.0	1.0	2.0	0.0	1.0	0.0	0.0	0.0
511	1	1.0	2.0	1.0	2.0	0.0	0.0	1.0	1.0	0.0

512 rows × 31 columns

Rescaling

Reason:

- Data too large or too small will affect accuracy

Method:

- StandardScaler

Image Of Data Pre rescaling

	age	wives	children	childEdu	farmNum	farmSize	capitalInput	distance	laborDays
0	4.007333	2.0	11.0	0.0	2.0	4.0	3050.000000	9.00000	5.000000
1	3.713572	2.0	7.0	0.0	1.0	4.0	2000.000000	9.00000	5.000000
2	4.060443	2.0	9.0	0.0	3.0	4.0	1890.000000	9.00000	5.000000
3	3.806663	0.0	8.0	0.0	2.0	3.0	1520.000000	8.00000	3.000000
4	4.007333	0.0	9.0	0.0	1.0	1.0	900.000000	8.50000	3.000000
...	...	...	...	...	...	...	...	...	...
507	4.248495	1.0	6.0	0.0	2.0	2.0	4391.909871	5.45968	3.852459
508	4.043051	1.0	4.0	0.0	1.0	14.0	4391.909871	5.45968	3.852459
509	4.110874	1.0	4.0	0.0	1.0	3.0	4391.909871	5.45968	3.852459
510	4.025352	1.0	7.0	1.0	1.0	6.0	4391.909871	5.45968	3.852459
511	3.555348	1.0	4.0	4.0	2.0	6.0	4391.909871	5.45968	3.852459

512 rows × 9 columns

Image Of Data POST rescaling

	age	wives	children	childedu	farmNum	farmSize	capitalInput	distance	laborDays
0	0.617177	2.485028	2.278734	-1.097872	-0.120875	-0.301513	-0.305989	7.807920e-01	8.766173e-01
1	-0.565138	2.485028	0.821996	-1.097872	-0.731231	-0.301513	-0.545416	7.807920e-01	8.766173e-01
2	0.830929	2.485028	1.550365	-1.097872	0.489482	-0.301513	-0.570499	7.807920e-01	8.766173e-01
3	-0.190472	-0.991295	1.186180	-1.097872	-0.120875	-0.440131	-0.654868	5.602492e-01	-6.512014e-01
4	0.617177	-0.991295	1.550365	-1.097872	-0.731231	-0.717366	-0.796244	6.705206e-01	-6.512014e-01
...	...	...	...	...	...	...	...	...	...
507	1.587792	0.746866	0.457811	-1.097872	-0.120875	-0.578748	0.000000	-7.835252e-16	-4.553255e-08
508	0.760932	0.746866	-0.270558	-1.097872	-0.731231	1.084661	0.000000	-7.835252e-16	-4.553255e-08
509	1.033900	0.746866	-0.270558	-1.097872	-0.731231	-0.440131	0.000000	-7.835252e-16	-4.553255e-08
510	0.689695	0.746866	0.821996	-0.625907	-0.731231	-0.024278	0.000000	-7.835252e-16	-4.553255e-08
511	-1.201949	0.746866	-0.270558	0.789989	-0.120875	-0.024278	0.000000	-7.835252e-16	-4.553255e-08

512 rows × 9 columns

Feature Engineering

Reasons:

- Make variable more relevant

Method:

- Data transformation(log)
- Interaction(product, division)

List of features you engineered

- Age ---> log(age)
- Age * Sex ---> Age_sex
- Number of children in education ---> Number of children in edu/ number of children

Step 4:
Train and Test Models

Models

Simple Linear regression baseline model

Lasso Regression round two ‘variable selection’

Ridge Regression using re-selected variables

Polynomial ridge regression more complicated ridge by polynomial feature **Engineering**

Kernel ridge regression more developed ridge with feature engineered

Support Vector regression same feature engineering with different loss function

Decision Tree regression classic model

Random Forest Regression (ensemble of tree model)

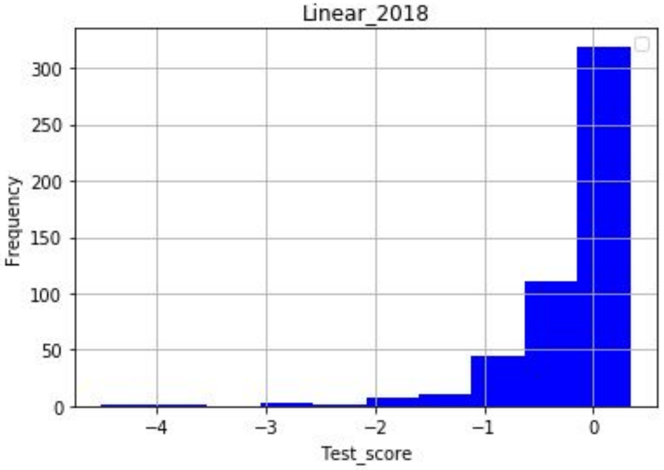
Simple Linear Regression

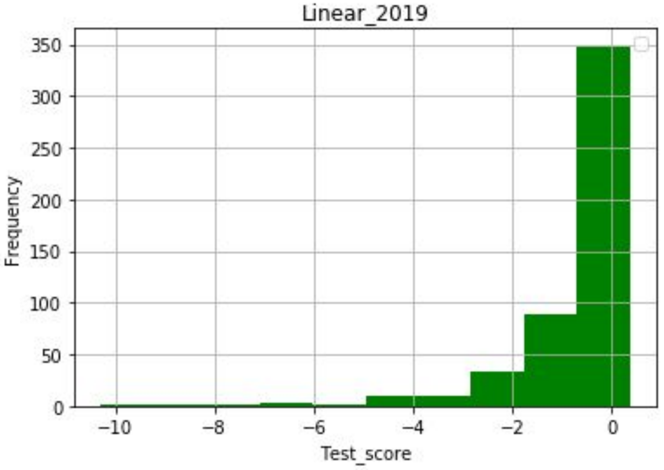
- **Pros:**
  - Simple method
  - Good interpretation
  - Easy to implement

- **Cons:**
  - Assumes linear relationship between dependent and independent variables, which is incorrect in most cases
  - Sensitive to outliers
  - If the number of observations are less, it leads to overfitting, it starts considering noise.

Simple Regression Performance

Target year	Test score (R_square)	Mean Square error	Mean Absolute error
2018	-0.170130244146	4456.308988872526	30.32341870526361
2019	-0.8009841908073049	162.1933074400931	5.732998098786998



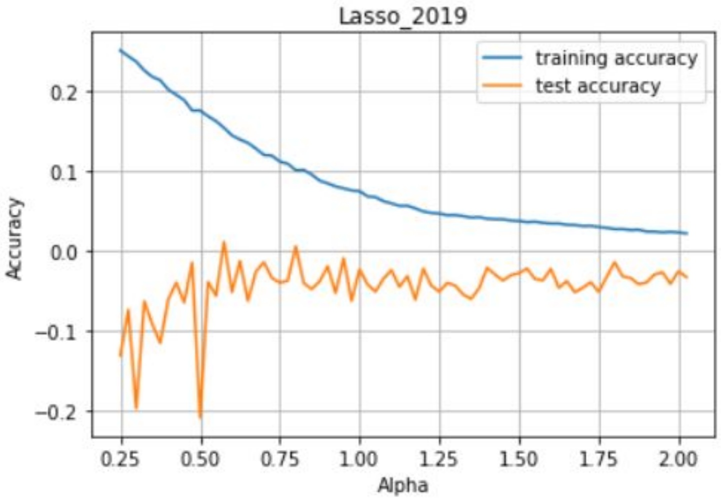


Lasso Regression

- Pros
  - Select features, by shrinking co-efficient towards zero.
  - Avoids overfitting
- Cons
  - Selected features will be highly biased.
  - LASSO will select only one feature from a group of correlated features, the selection is arbitrary in nature.
  - For different bootstrapped data, the feature selected can be very different.
  - Prediction performance is worse than Ridge regression.

Lasso Regression Performance

Target year	Test score (R_square)	Mean Square error	Mean Absolute error
2018	0.11548469188164179	3642.504229413452	24.27636432597194
2019	0.01175960127184189	119.55704347681979	3.9179119891874743

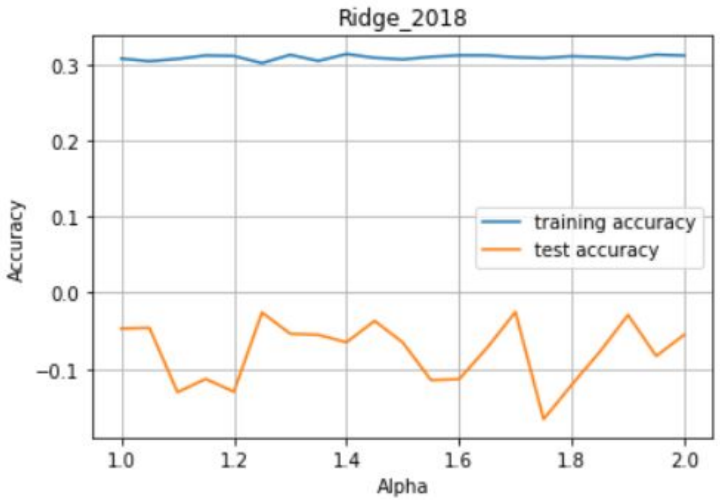


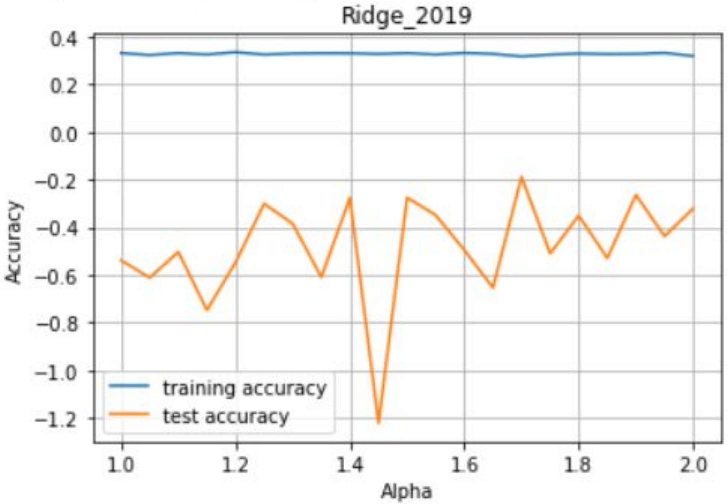
Ridge Regression

- **Pros**
  - Trades variance for bias (i.e. in presence of collinearity, it is worth to have biased results, in order to lower the variance.)
  - Prevents overfitting
- **Cons**
  - Increases bias
  - Need to select perfect alpha (hyper parameter)
  - Model interpret-ability is low

Ridge Regression Performance

Target year	Test score (R_square)	Mean Square error	Mean Absolute error
2018	-0.02518524395712685	3397.314804511116	27.035674656887366
2019	-0.18782652218740098	129.23338586848988	4.999598422288299



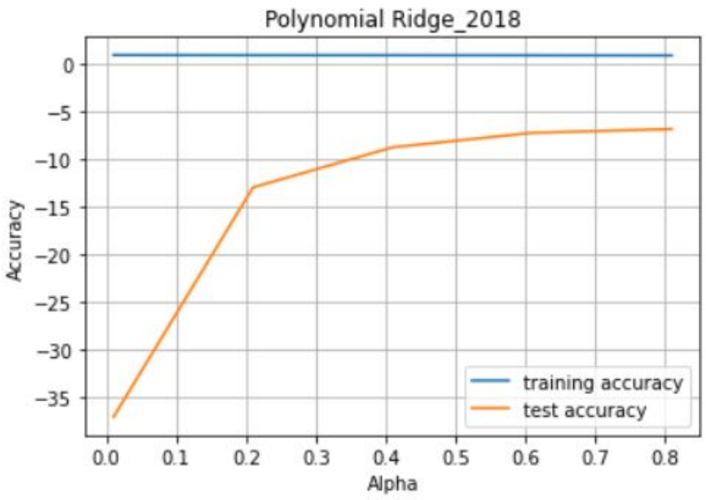


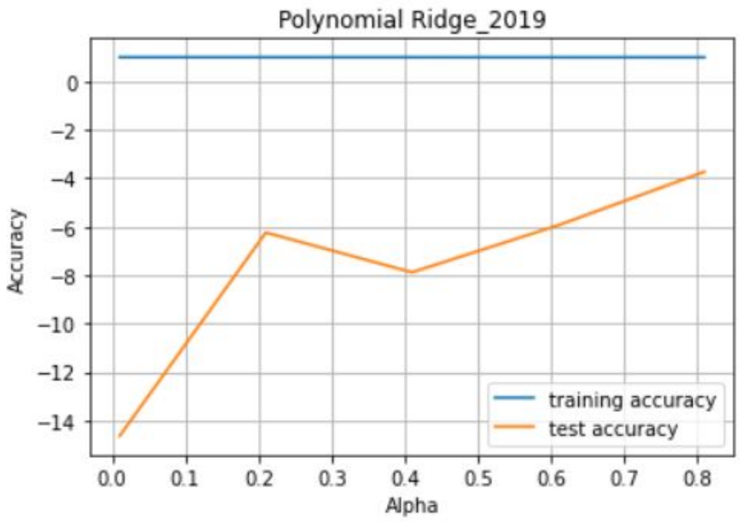
Polynomial Ridge Model

- Pros
  - Works on any size of the dataset
  - Works very well on non-linear problems
  - The Ridge parameter is to prevent overfitting.
- Cons
  - We need to choose the right polynomial degree for good bias/variance tradeoff

Polynomial Ridge Model Performance

Target year	Test score (R_square)	Mean Square error	Mean Absolute error
2018	-6.790067973302218	16164.718542838154	58.803112331283764
2019	-3.7296243659363197	365.35777649745614	8.377060862384722





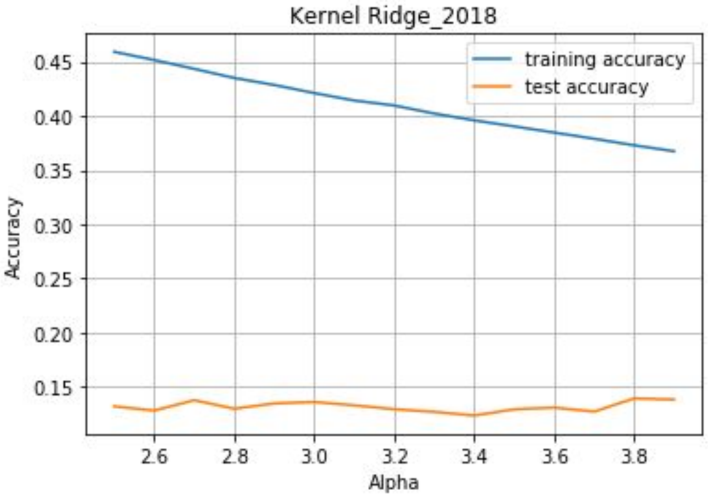
Kernal Ridge Model

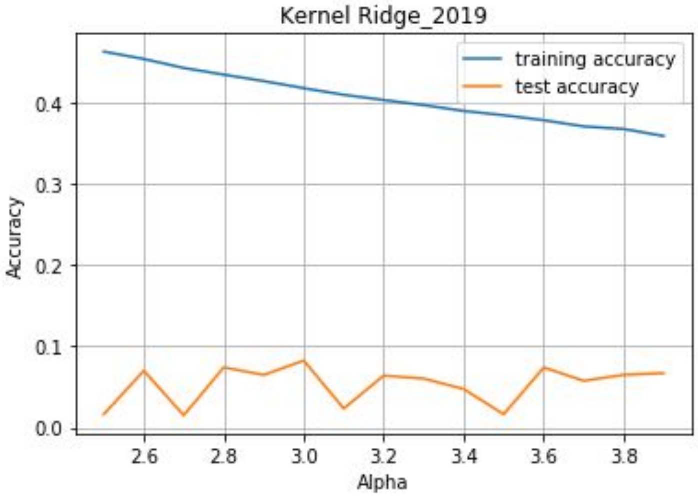
- Pros
  - Typically faster for medium-sized datasets
  - Regularization of overfitting

- Cons
  - May have scaler issue when predict time series

Kernel Ridge Regression Performance

Target year	Test score (R_square)	Mean Square error	Mean Absolute error
2018	0.147778169351074	3519.1063911862257	8.377060862384722
2019	0.09736766676515778	122.4926899647815	3.268414828384891



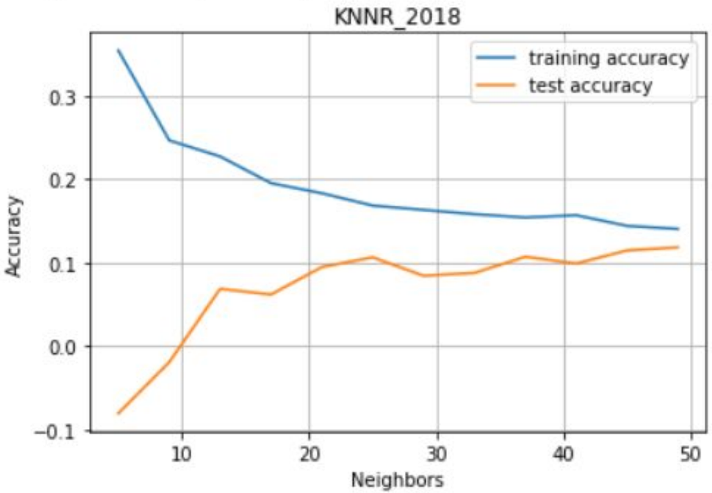


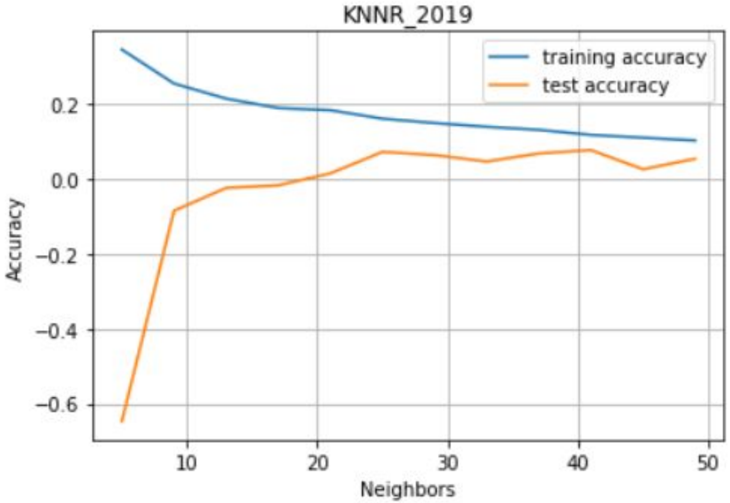
K Neighbors Regression

- Pros
  - Fairly intuitive and simple
  - No assumptions required
  - New data can be added seamlessly, which will not impact the accuracy of the algorithm
- Cons
  - Does not work well with high dimensions
  - Sensitive to noisy data, missing values and outliers
  - Does not work well with large data sets, as the cost of calculating distance is huge

K Neighbors Regression Performance

Target year	Test score (R_square)	Mean Square error	Mean Absolute error
2018	0.1184961389067918	3623.0327645164084	22.240634531546085
2019	0.07808626299422096	124.71995910419524	3.4785359651478376



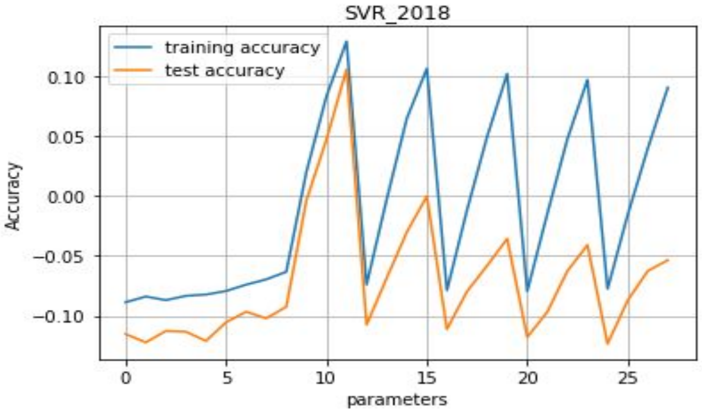


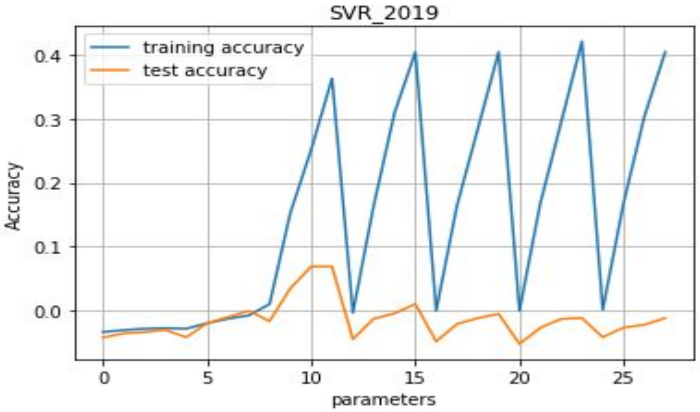
Support Vector Regression

- Pros
  - Easily adaptable
  - Works very well on non-linear problems
  - Not biased by outliers
- Cons
  - Compulsory to apply feature scaling
  - Difficult to understand

Support Vector Regression Performance

Target year	Test score (R_square)	Mean Square error	Mean Absolute error
2018	0.08405774507202138	4120.783386330159	19.3736347954319
2019	0.0713225362497389	126.02479108054344	2.58752362022368 8



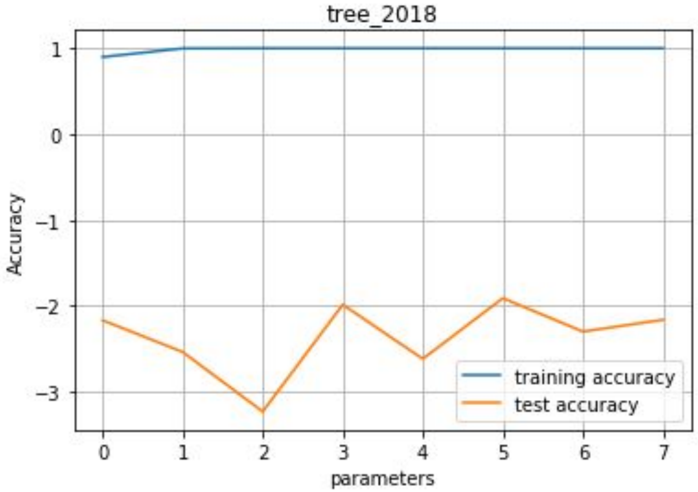


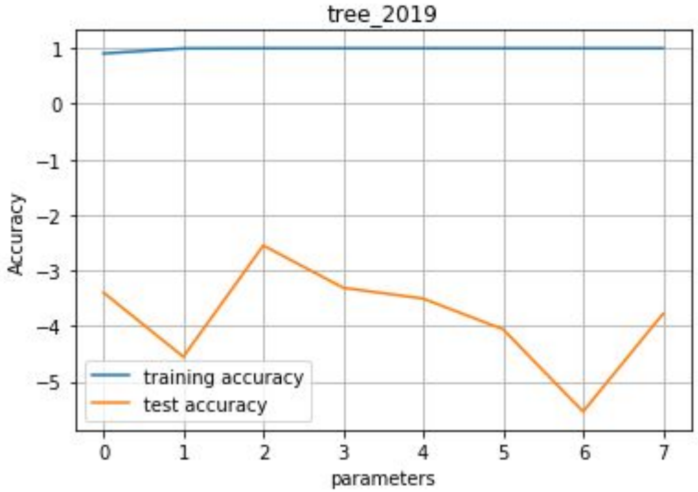
Decision Tree Regression

- Pros
  - Easily handles both discrete and continuous variables
  - Ignores irrelevant information
  - Fast predictions
  - Does not require standardization and normalization
- Cons
  - Fitting can be mysterious (instabilities)
  - Can easily overfit, due to greedy strategy
  - Higher time to train the model

Decision Tree Regression Performance

Target year	Test score (R_square)	Mean Square error	Mean Absolute error
2018	-1.7985938865192725	7460.2690258156435	25.985524888033424
2019	-2.0343886154499127	289.76469254753624	5.032419455887545





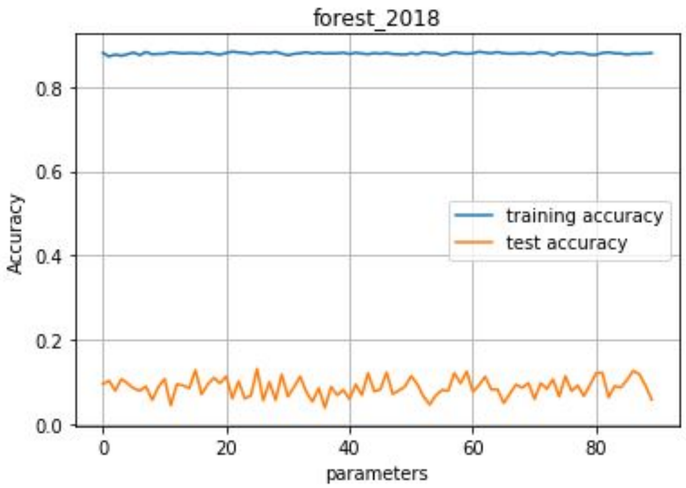
Random Forest Regression

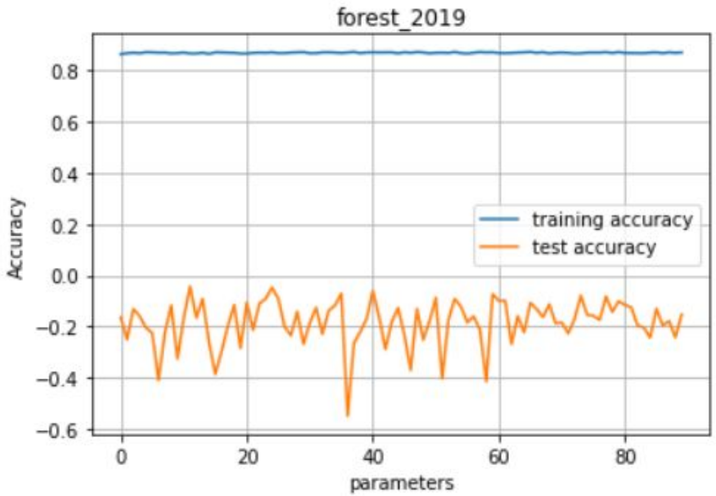
- **Pros:**
    - One third of data is not used for training, hence it can be used for testing.
    - High performance and accurate

- **Cons:**
    - Less interpret-ability, black box approach
    - Can over fit the data.
    - Requires more computational resources
    - Prediction time is high

Random Forest Regression Performance

Target year	Test score (R_square)	Mean Square error	Mean Absolute error
2018	0.16612694324909824	2965.9010923773585	20.598896194045743
2019	-0.05257094440610835	110.11972830905844	3.6916589605892205





Step 5:
Model Comparison

Model Comparison

Model	Score		MSE		MAE	
	2018	2019	2018	2019	2018	2019
Simple Linear Regression	-0.17013	-0.80098	4456.30899	162.19331	30.32342	5.73300
Ridge Regression	-0.02519	-0.18783	3397.31480	129.23339	27.03567	4.99960
Lasso Regression	0.11548	0.01176	3642.50423	119.55704	24.27636	3.91791
Polynomial Ridge Model	-6.79007	-3.72962	16164.71854	365.35778	58.80311	8.37706
Kernel Ridge Regression	0.14778	0.09737	3519.10639	122.49269	8.37706	3.26841
Support vector regression	0.08406	0.07132	4120.78339	126.02479	19.37363	2.58752
K Neighbors Regression	0.11850	0.07809	3623.03276	124.71996	22.24063	3.47854
Decision Tree Regression	-1.79859	-2.03439	7460.26903	289.76469	25.98552	5.03242
Random Forest Regression	0.16613	-0.05257	2965.90109	110.11973	20.59890	3.69166

Insights and Implications

Insights

Model

1. Kernel Ridge regression is the best method, while polynomial with 2 degree is bad, implying the subtle and complicated relationship does exist between the chosen X variables and the target variable.
2. K_neighbor regression and Support vector machine method can be a competing choice in application, but not as explanatory and easy to understand, also with a low speed.

Data:

Cleaner data will define better and more accurate models !

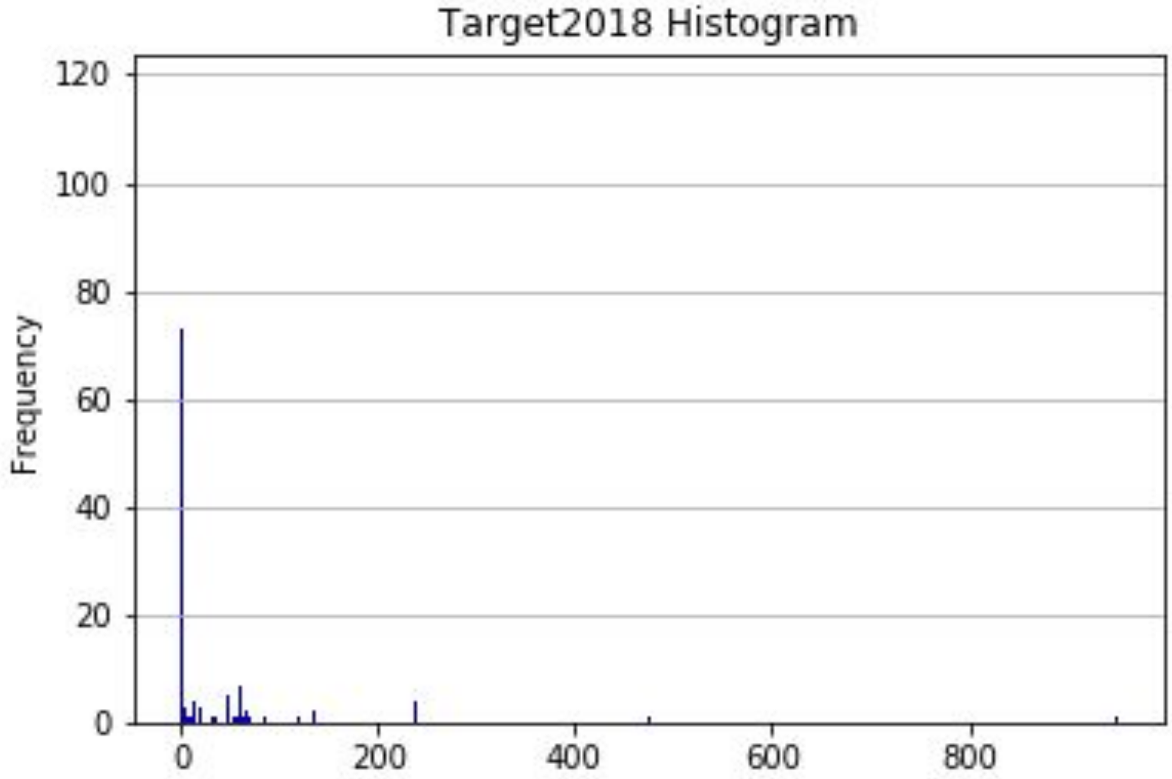
Which features matter the most?

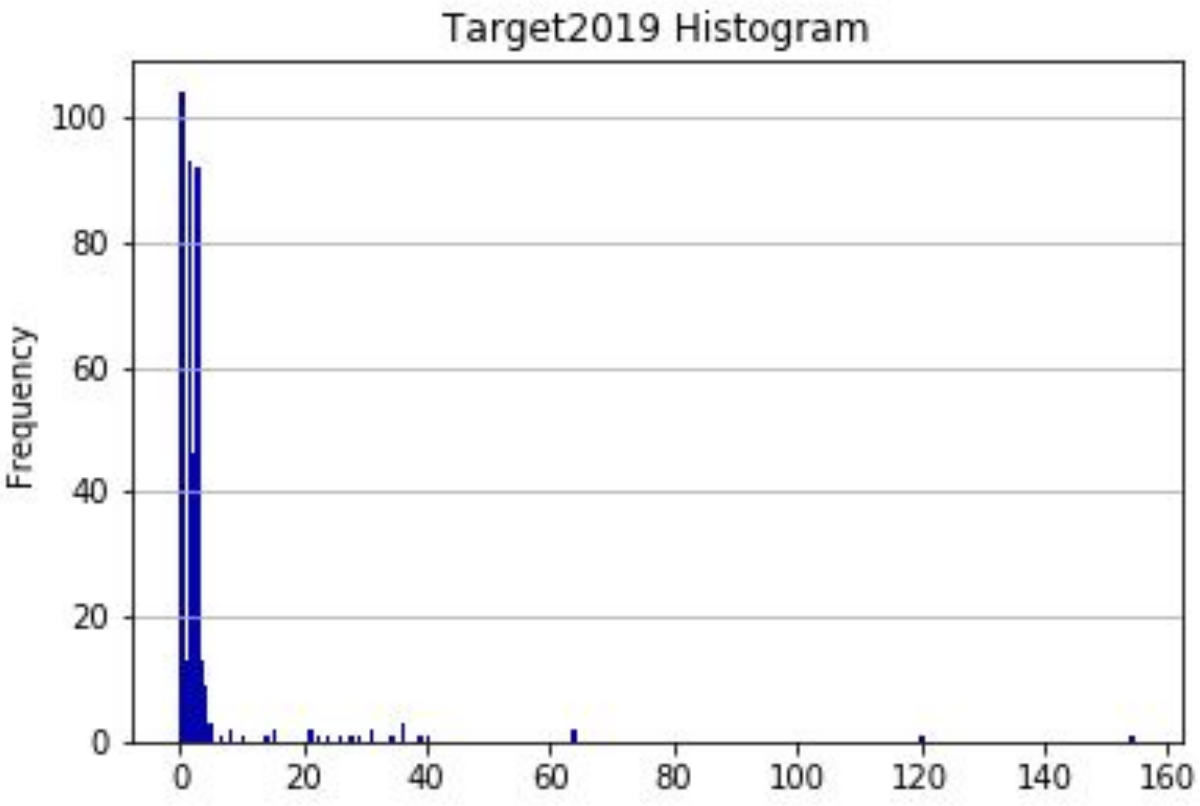
Top 10 with the highest predicted profit rate:
(Here are some of the features with relatively high similarity among these selected people)

Name of Farmer	Age	Sex	Head_household	Marriage	# of children	Education level	Distance	Labor_days
Clement Attakora	48	Male	1	1	6	JHS	6km	5
TWUM KWAME	50	MALE	1	1	6	2	4 MILES	5
BOATENG JOSEPH	67	MALE	1	1	7	2	1 MILE	6
Yaro Ndannai	45	Male	1	1	6	J.S.S	6km	5
ADJEI BAAH	62	MALE	1	1	6	2	5 MILES	5
COMFORT OKYERE	62	FEMALE	0	1	4	2	6 MILES	7
ANTWI BOASIAKO	78	MALE	1	1	8	1	1 MILE	4
BOSOMPEM DANIEL	46	MALE	1	1	6	1	2 MILES	3
HAKEEM MARFO	48	MALE	1	1	6	1	6 MILES	5
MARGARET BOAHEN	62	FEMALE	1	0	7	1	0.5 MILES	3

Q & A

Appendix


Target2018 Histogram
Frequency
0
20
40
60
80
100
120
0
200
400
600
800



Target2019 Histogram
Frequency
100
80
60
40
20
0
0
20
40
60
80
100
120
140
160

Disregarded Variables

Category	Disregarded Values
Farmers' Info	Relationship to head of house Number of children in higher education
Farms condition & Crops Info & Livestock	Main_crop Farm_category Agri_type breed
Disaster & Disease	Type of pest Type of disease?
Irrigate & Soil management	irrigate_type Are you using soil management/ fertilization If fertilizer what type
Info & Financing	History in receiving extension services was extension services relevant Access_date Target_main target_other

Variables not available in dataset

Category	VARIABLES - Theoretical	VARIABLES - Possible
Productivity	**Mechanization level**	What kind of farming techniques been used? /what kind of automated machinery been applied? /If so how many?
	Human Resource	Family structure, how many adults(men and women) in the family, and how many children? / How many labor and non-labor in the family?
		Does he/she hire additional people from outside the family as labors ?
	Education or knowledge in farming tech	Have you accepted education or training in farming ? / What kind of farming techniques been used?
Storage and transportation	**Storage method/ Transportation instrument**	What's the specific storage approach? What kind of/ how many instruments be used to do transportation/ how many cars owned?
Capital resource	**Government Subsidy**	Yes/no? If yes, how much?
	Houses and other fixed assets ?	How many houses/other fixed assets owned?

Data Audit

Independent Variables:

- Farmer’s info
- Farm’s info
- Operation mgmt info
- External environment condition
- Marketing efficiency

Dependent Variables:

- Yield
- Sales
- Price
- Operation Cost

Numerical Data

	age	wives	children	childedu	farmNum	farmSize	capitalInput	distance	laborDays
0	0.617177	2.485028	2.278734	-1.097872	-0.120875	-0.301513	-0.305989	7.807920e-01	8.766173e-01
1	-0.565138	2.485028	0.821996	-1.097872	-0.731231	-0.301513	-0.545416	7.807920e-01	8.766173e-01
2	0.830929	2.485028	1.550365	-1.097872	0.489482	-0.301513	-0.570499	7.807920e-01	8.766173e-01
3	-0.190472	-0.991295	1.186180	-1.097872	-0.120875	-0.440131	-0.654868	5.602492e-01	-6.512014e-01
4	0.617177	-0.991295	1.550365	-1.097872	-0.731231	-0.717366	-0.796244	6.705206e-01	-6.512014e-01
...	...	...	...	...	...	...	...	...	...
507	1.587792	0.746866	0.457811	-1.097872	-0.120875	-0.578748	0.000000	-7.835252e-16	-4.553255e-08
508	0.760932	0.746866	-0.270558	-1.097872	-0.731231	1.084661	0.000000	-7.835252e-16	-4.553255e-08
509	1.033900	0.746866	-0.270558	-1.097872	-0.731231	-0.440131	0.000000	-7.835252e-16	-4.553255e-08
510	0.689695	0.746866	0.821996	-0.625907	-0.731231	-0.024278	0.000000	-7.835252e-16	-4.553255e-08
511	-1.201949	0.746866	-0.270558	0.789989	-0.120875	-0.024278	0.000000	-7.835252e-16	-4.553255e-08

512 rows × 9 columns

Category Data

	household	marriage	farmerEdu	farmPrimary	farmType	otherIncome	drought	fires	flood	wind
0	1	1.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0
1	1	1.0	1.0	1.0	1.0	1.0	0.0	0.0	1.0	0.0
2	1	1.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0
3	1	1.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0
4	1	1.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...
507	0	1.0	2.0	1.0	2.0	1.0	0.0	0.0	0.0	1.0
508	1	1.0	2.0	1.0	2.0	0.0	1.0	0.0	0.0	0.0
509	0	1.0	2.0	1.0	2.0	0.0	0.0	0.0	0.0	0.0
510	1	1.0	0.0	1.0	2.0	0.0	1.0	0.0	0.0	0.0
511	1	1.0	2.0	1.0	2.0	0.0	0.0	1.0	1.0	0.0

512 rows × 31 columns