CattleGo: An Interactive Platform for US Cattle Production Visualization and Prediction

tgao68, xcao67, qshao9, czhang487, yliu3093, yzhao464
CSE6242 Data & Visual Analytics
Georgia Institute of Technology

Summary

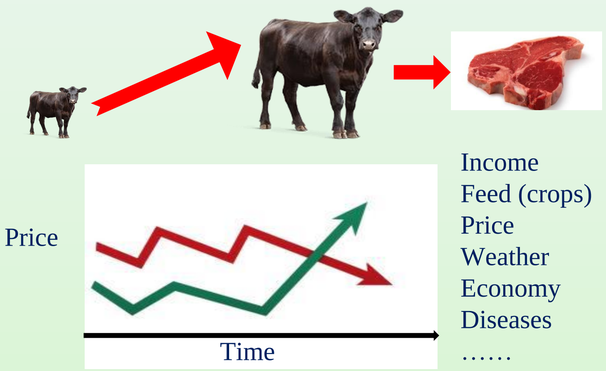
The U.S. cattle industry is a $67.56 billion/year business and a major contributor to the global food supply. Thus, it is of great importance to maintain an organic growth of the U.S. cattle industry.

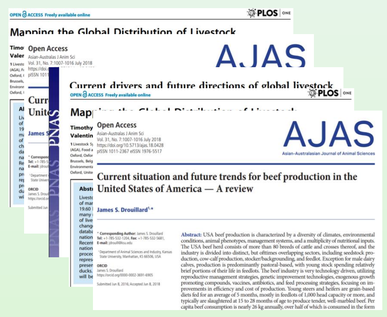
Due to the nature of this business, cattle farmers are especially vulnerable to price fluctuations in the market. Therefore, it is critical for farmers to have easy access to cattle production data to project future market demands and maximize their interest.

To help farmers, we built an interactive platform to provide geographical and statistical representations of historical cattle production, to pinpoint the key factors influencing the cattle production industry.

The current problem

Cattle production is a multi-year practice, farmers need to plan many things ahead of time, and the price of beef is sensitive to many factors. If the farmers can not foresee market demands in the future, they might risk their whole operation. Therefore, it is critical for farmers to have easy access to cattle production data to project future market demands. However, currently, cattle production data and prediction is not easily accessible. Besides, the process is hard for most farmers.





Information is not easily accessible.

Aim

We aim to develop an interactive web app as intuitive as possible, especially for non-data-oriented users. We want the web app to help users have a clear view of the changes in cattle production over the years and reveal relationships between the cattle production data and some important factors.

Data

Temporal data were collected from the databases of National Agricultural Statistics Service, Bureau of the Census and National Oceanic and Atmospheric Administration.

Agricultural data (41 million records, 737MB)

Weather data (78920 CSV files, 637MB)

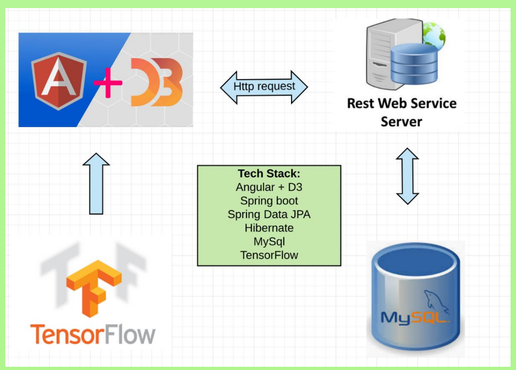
Filter, Clean, Merge

Year	Cattle production in lb	Corn production in tons	Soybeans production in bu	STATE	TMAX	TMIN	TAVG	EMXT	EMNT	DX90	DT32	PRCP	SNOW	AWND
1988	671,520,000	160,000	14250000	AL	23.512941	10.138222	16.757727	37.796078	-10.46	67.568627	63.2	1286.1941	19.87619	3.3
1989	680,350,000	250,000	11970000	AL	22.942778	10.65	16.725833	35.744444	-16.494	44.814815	50.36	1635.6486	4.7663551	3.22
1990	641,115,000	200,000	7480000	AL	25.110615	11.219516	18.175161	38.084615	-7.325806	84.2	38.048387	1443.0634	0.024	3.34
1991	659,250,000	350,000	8050000	AL	23.856379	11.57963	17.725472	36.210345	-9.618519	60.551724	50.574074	1594.3113	1.5952381	3.2
1992	525,310,000	325,000	7830000	AL	22.826	10.168	16.4546	35.670909	-8.516	34.636364	51.78	1496.424	20.207207	3.14
1993	508,476,000	225,000	7080000	AL	23.144762	10.2685	16.697119	37.211111	-10.035	72.126984	55.6	1275.0531	106.99091	3.2

Final data: 1502 records for each state from 1988 to 2017

Architecture Design

The core components are: the REST Web Service Server developed by Spring boot framework, MySql database, Machine learning component by TensorFlow and a client application by Angular and D3.

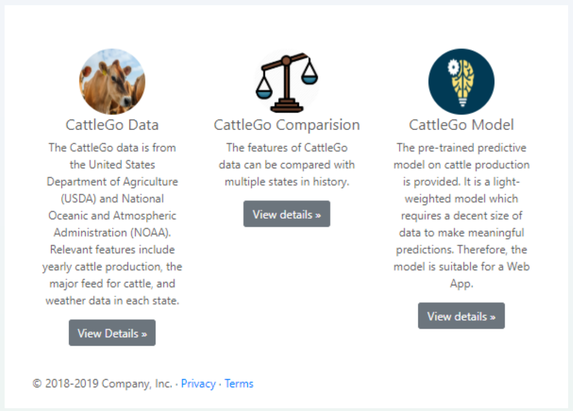


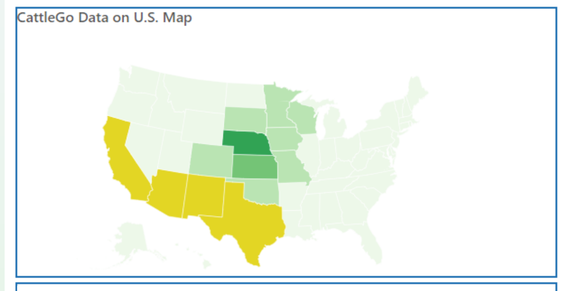
Angular provides the platform for our web application.

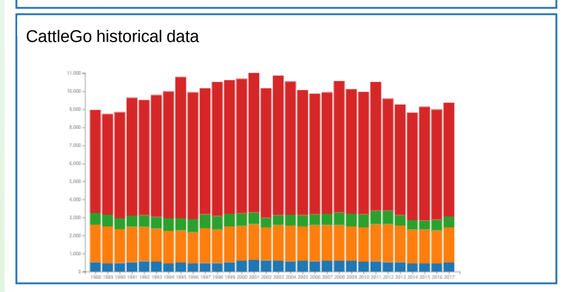
TensorFlow is used to develop and train machine learning models.

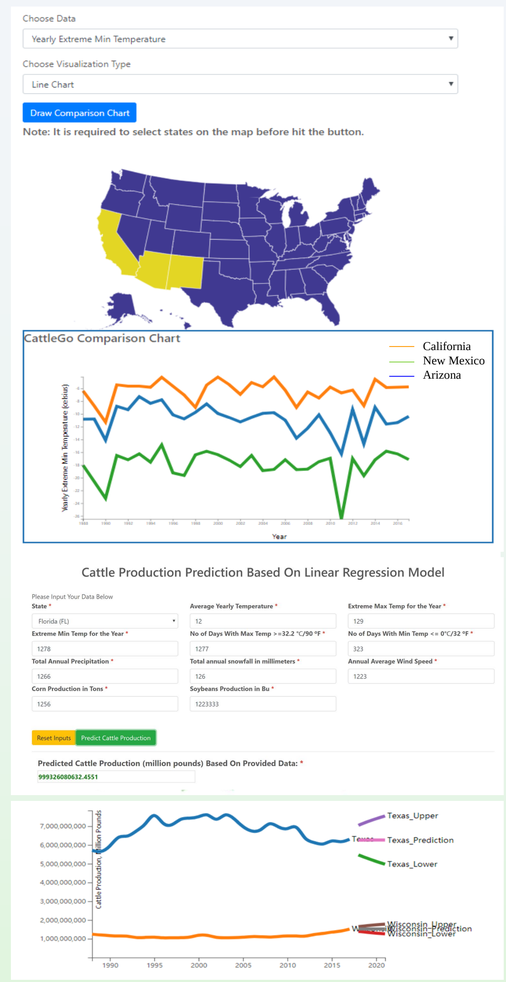
D3 enables dynamic, interactive data visualizations in web browsers.

Preview of CattleGO



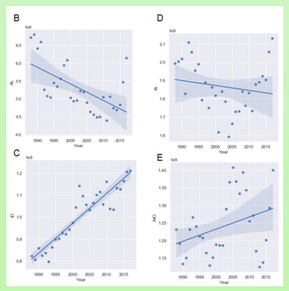


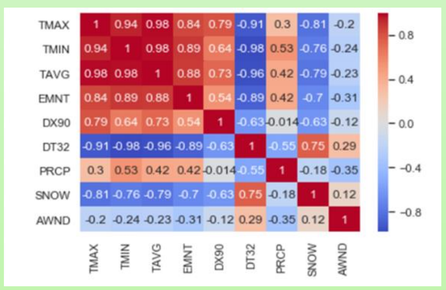


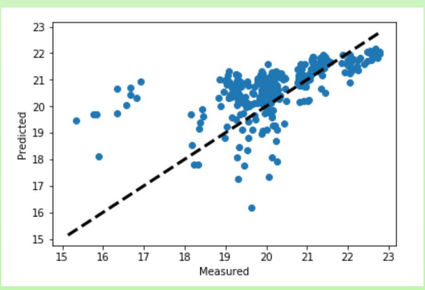


1. **CattleGO Data**: users can choose the data category and year. After selection, an US-map based heatmap will be generated. Users can hover the mouse to see the details of a specific state.
2. **CattleGO Comparison**: users can choose between stack bar and line chart for data presentation. Data from different states can be viewed at the same time for better comparison.
3. **CattleGO Model**: both time series and regression models have been implemented to predict the production of cattle. Users can access these model by clicking the CattleGo Model button.

Results



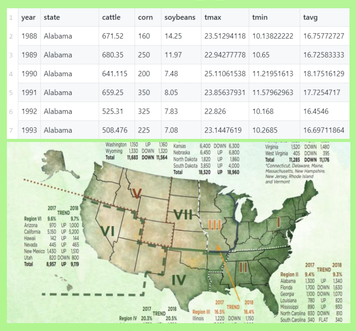




1. Cattle production fluctuates with different patterns for different states over the years.
2. Weather features are linearly correlated with each other, indicating possible multicollinearity. Thus LASSO regression was used to perform variable selection.
3. Different machine learning models were explored. Linear regression can achieve R^2 of 0.60 on the test data. ARIMA models provides prediction for the next 4 years with 95% confidence interval.

The Comparison

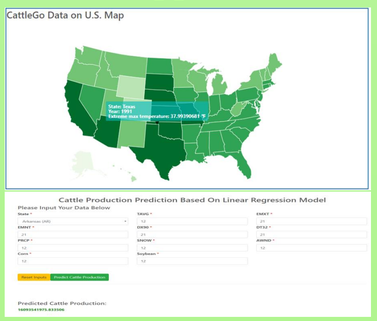
Previous approach



- Line by line numbers
- Static Figures
- Not intuitive
- Not easily accessible

VS

Our approach



- Interactive
- Intuitive
- Easily accessible
- Predictions available