Bio-Tesseract

Optimizing Production at Chinese Biogas Plants

TEAM 2

AKSHAT KHANDELWAL
ASHIS GHOSH
POL MOLINAS
ROHAN KULKARNI
SHILIN CHEN

project overview.

biogas in China.

1. China is the second largest energy consumer
2. Produces massive industrial organic waste
3. Resource per capita in shortage

potential users.

PLANT MANAGERS maximize biogas output and minimize auxiliary inputs and undesired outputs.

NEW PLANTS need for assessing the feasibility of new biogas plant.

approach.

HAINAN DATA

- Day level data
- 30 columFrom 05/01/14 to 02/28/17 (1,035 days)

DATA CLEANING

- Drop unused columns.
- Rename column titles.
- Convert strings to numbers
- Fill NaN values with 0.

FEATURE ENGINEERING

1. Carbon/hydrogen ratio composition
2. **Combining inputs based on mass input**

DATA MANIPULATION

- **Time shifting**
- **Combining** rows of data **cumulatively** and on **rolling sum** basis

ML MODELS

- **Classification** (XG Boost, RF, DT, LogReg)
- **Linear Models** (LinReg, Ridge, Lasso)
- **Neural Networks**

BEST MODEL

Classification

- **Random forest**, time shift **21 days**, rolling sum of **40 days**, **10** buckets: **96.7%**

Regression Model

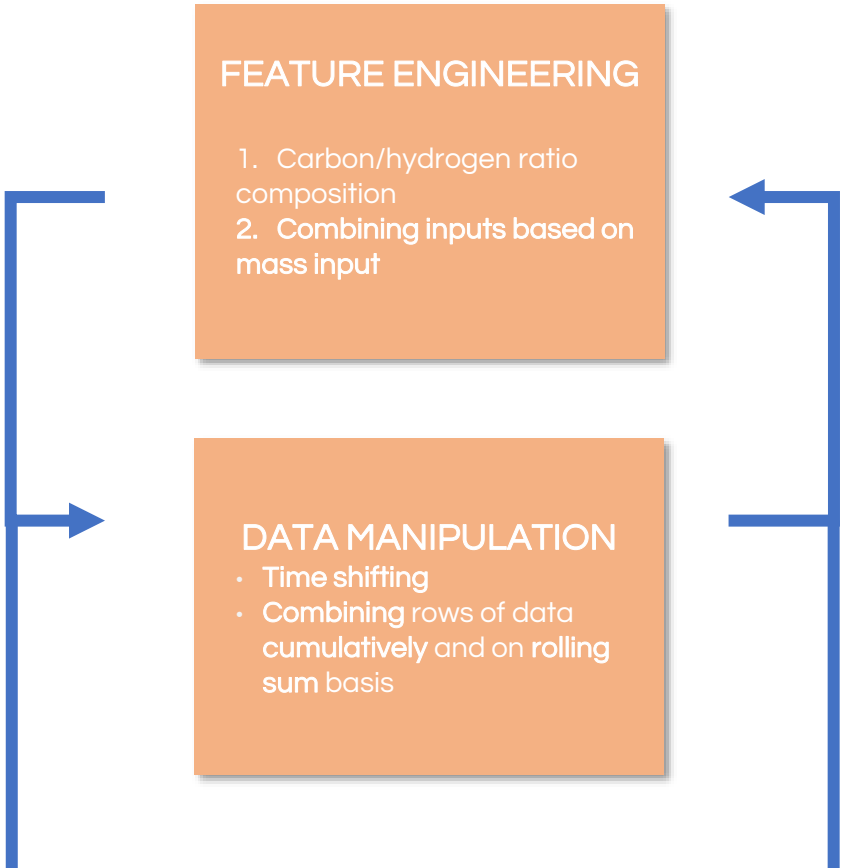
- **Linear Regression**, time shift **29 days**, rolling sum of **40 days**: **97.5%**

GUI

approach.

INPUTS

Pig Manure
Bagasse
Cassava
Waste Water
Kitchen Waste
Fecal Waste
Tea Waste
Chicken Waste
Alcowaste
Medicine Waste
Energy Grass
Banana Waste
Lemon Waste
Percolate
Other



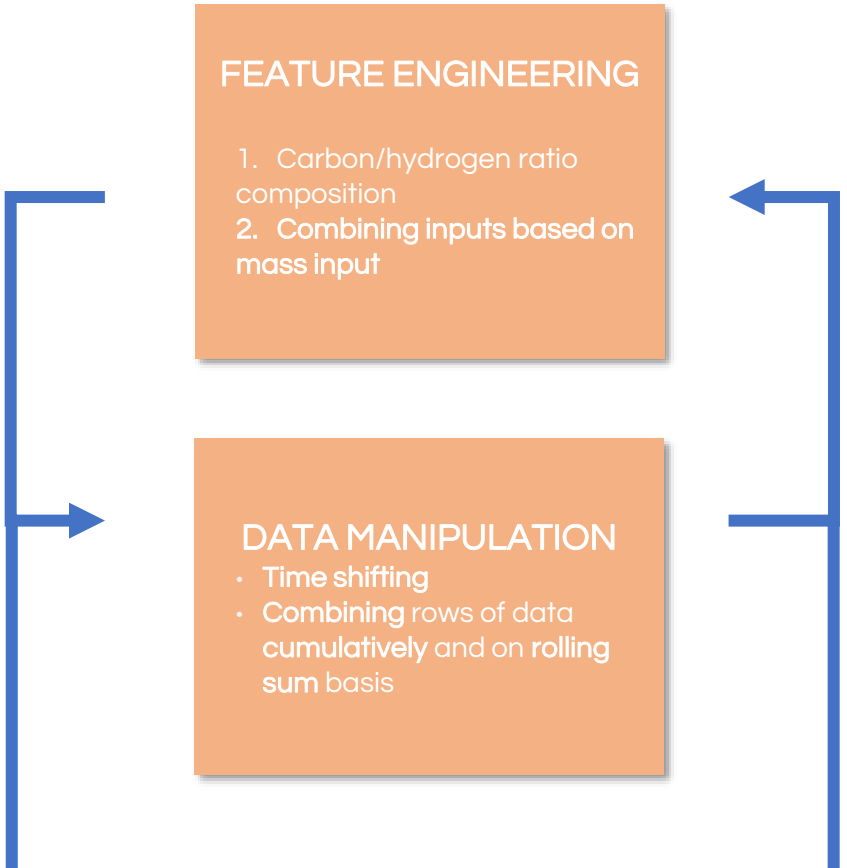
approach.

INPUTS

Pig Manure
Bagasse
Cassava
Waste Water
Kitchen Waste
Fecal Waste
Tea Waste
Chicken Waste
Alcowaste
Medicine Waste
Energy Grass
Banana Waste
Lemon Waste
Percolate
Other

Group 1

Group 2



approach.

INPUTS

Pig Manure
Bagasse
Cassava

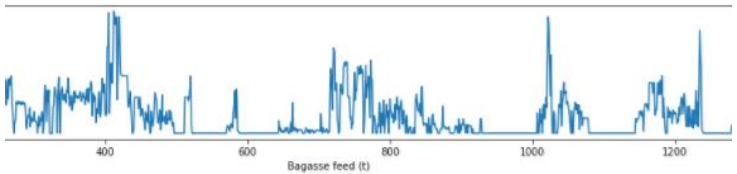
OUTPUT

Raw BioGas

time shifting.

Thermal stage	Process temperatures	Minimum retention time
psychrophilic	< 20 °C	70 to 80 days
mesophilic	30 to 42 °C	30 to 40 days
thermophilic	43 to 55 °C	15 to 20 days

rolling sum.



FEATURE ENGINEERING

1. Carbon/hydrogen ratio composition
2. **Combining inputs based on mass input**

DATA MANIPULATION

- **Time shifting**
- **Combining** rows of data **cumulatively** and on **rolling sum** basis

results.

- Time shifting
- Combining rows of data cumulatively and on rolling sum basis

ML MODELS

- **Classification** (XG Boost, RF, DT, LogReg)
- **Linear Models** (LinReg, Ridge, Lasso)
- **Neural Networks**

BEST MODEL

Classification

- **Random forest**, time shift **21 days**, rolling sum of **40 days**, **10** buckets: **96.7%**

Regression Model

- **Linear Regression**, time shift **29 days**, rolling sum of **40 days**: **97.5%**

results.

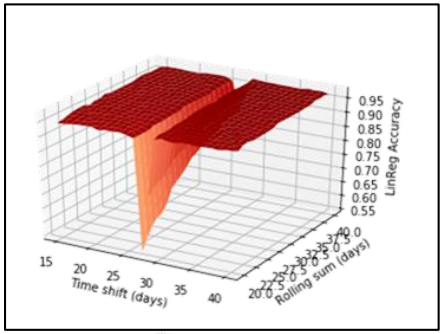
BEST MODEL

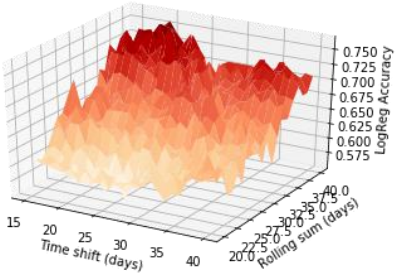
Classification

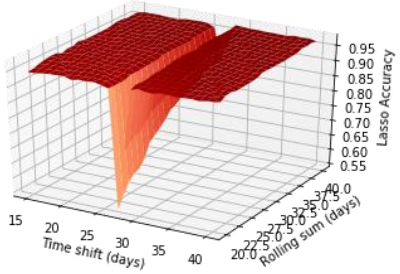
- **Random forest**, time shift **21 days**, rolling sum of **40 days**, **10** buckets: **96.7%**

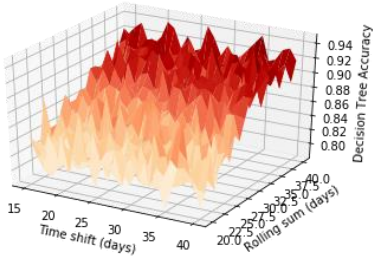
Regression Model

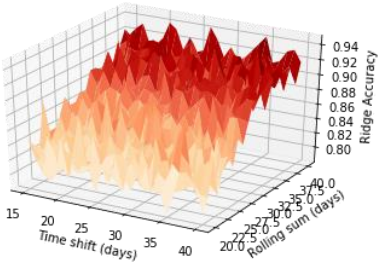
- **Linear Regression**, time shift **29 days**, rolling sum of **40 days**: **97.5%**

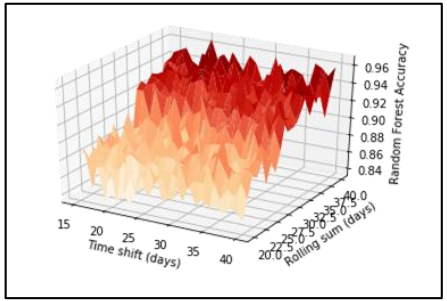












Approach + ui.

HAINAN DATA

- Day level data
- 30 columFrom 05/01/14 to 02/28/17 (1,035 days)

DATA CLEANING

- Drop unused columns.
- Rename column titles.
- Convert strings to numbers
- Fill NaN values with 0.

FEATURE ENGINEERING

1. Carbon/hydrogen ratio composition
2. **Combining inputs based on mass input**

DATA MANIPULATION

- **Time shifting**
- **Combining** rows of data **cumulatively** and on **rolling sum** basis

ML MODELS

- **Classification** (XG Boost, RF, DT, LogReg)
- **Linear Models** (LinReg, Ridge, Lasso)
- **Neural Networks**

BEST MODEL

Classification

- **Random forest**, time shift **21 days**, rolling sum of **40 days**, **10** buckets: **96.7%**

Regression Model

- **Linear Regression**, time shift **29 days**, rolling sum of **40 days**: **97.5%**

GUI

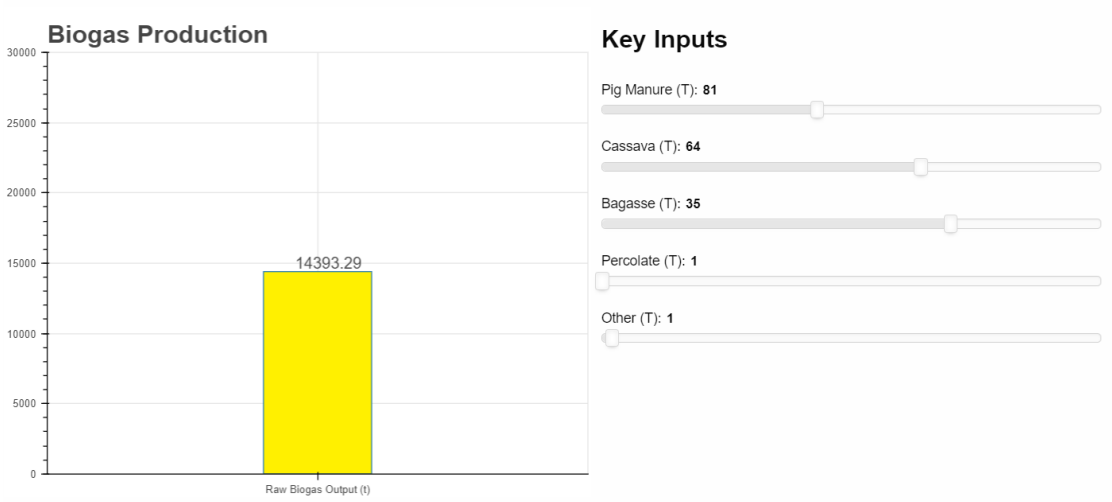
UI.

BEST MODEL

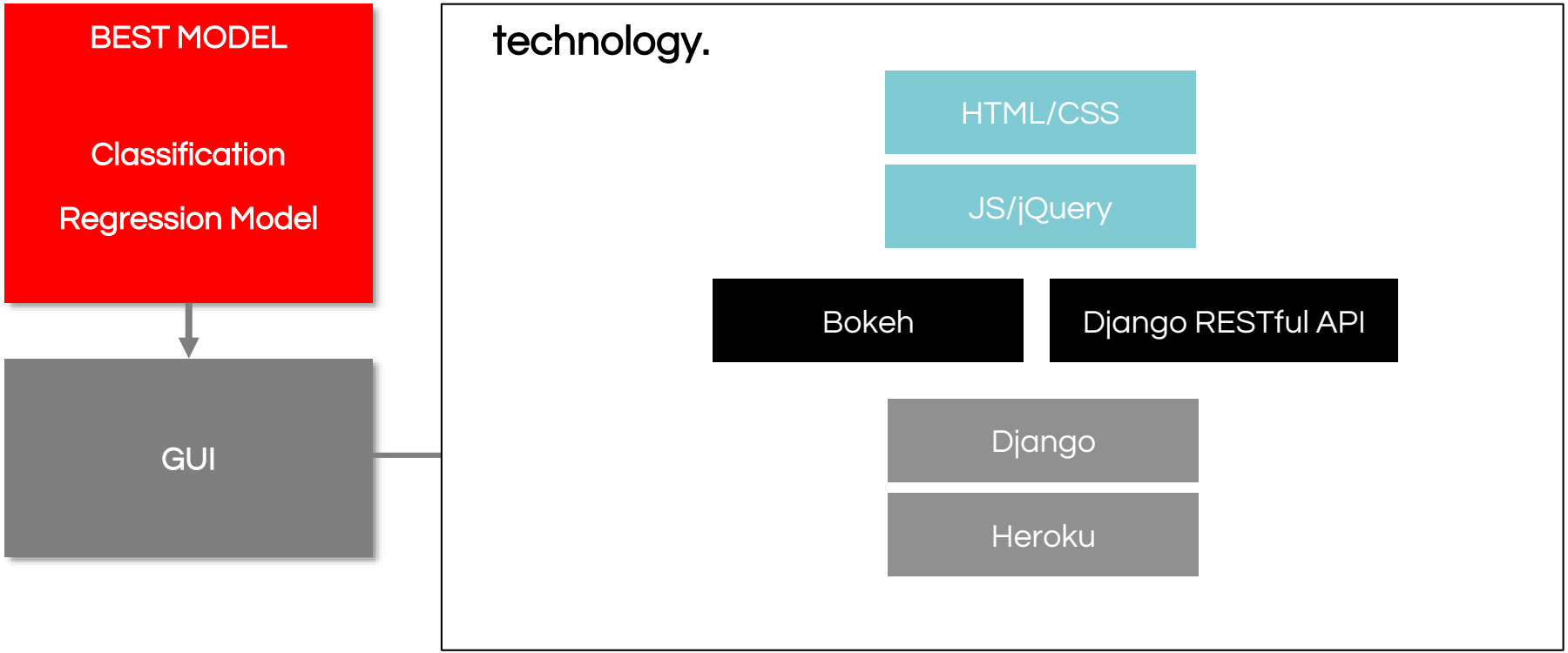
Classification

Regression Model

GUI



UI.


BEST MODEL
Classification
Regression Model
GUI
technology.
HTML/CSS
JS/jQuery
Bokeh
Django RESTful API
Django
Heroku

UI.

BEST MODEL

Classification

Regression Model

GUI

user flow.

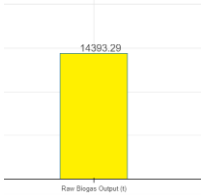
Input

Front End

Model API

Hosting/Backend

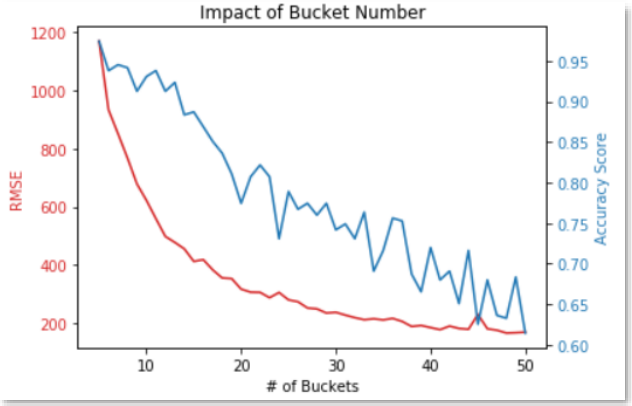
Visualization



UI demo.

main takeaways.

1. Classification models are not a good choice to model continuous data



2. Feature engineering has to have actual meaning for your data

3. It is important to work on one model at a time and move up as necessary

future work.

1. Optimization of the raw inputs to achieve maximum biogas outputs.
   → From a UX perspective – the plant manager can set their desired biogas output and see how they must adjust their inputs to achieve that

2. Model other outputs other than BioGas (such as fertilizer)

3. Include data from other plants for a more generalized model, or to develop models for various plant types

4. Modify UX/UI to allow for ‘time-based’ inputs.
   → Allow for material inputs to be input as a varying amount over a change of time

THANK YOU.

Q&A