Identifying Relationships between Customer Accounts

PURDUE UNIVERSITY | The Data Mine

Amogh Jani, Reis McMillan, Kevin Thierry Affokou, Noah Sakaitani, Raj Khera, Max Ruderman

CO-ALLIANCE

INTRODUCTION

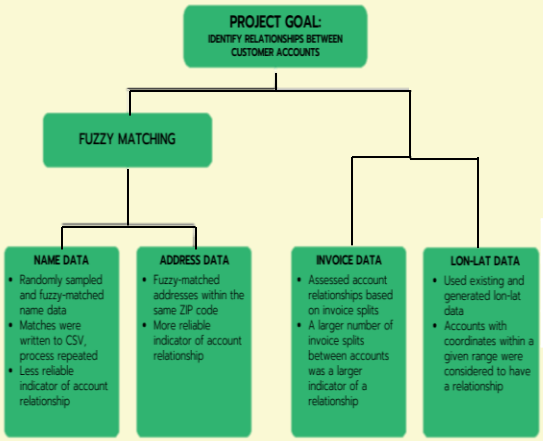
Company Intro:

- Co-Alliance: Cooperative founded in 2002; operates in Indiana, Illinois, Ohio, and Michigan.
- 4 main business units:
  - Agronomy, Energy, Grain, and Swine and Animal Nutrition
- One of the ten largest agriculture retailers in the US.

Objectives:

- Establish relationship between customers based on matching criteria (address, name and Latitude & Longitude).
- Create maps showing location of customers and Co-Alliance facilities.
- Build model to recommend customers for targeted divisional ad campaign.

Approach:



RESEARCH

Final Table

- **Merged Results of Algorithms to create list of GrowID pairs**
  - Latitude-Longitude distance, name, address, and invoices all create pairs of grower-ids that show a "match"

- **Duplicates were removed**
  - Pairs that matched on more than one algorithm were merged
- **Incorporated Agvend Data to the final table**
  - Agvend Data contains the pairs known to be matches
  - Added a boolean column if the pair shows up in the Agvend Data
  - Allows for the team to know which pairs are not yet in Agvend Data

Data Validation

Problems Found

- **Addresses - missing lat-long matches for people with the same addresses and some of the coordinates for the data**
  - using openstreetmap, we can now make maps to visualize which areas can better served

- **Invoices - missing some of the invoice split based on known splits**
  - reworked the code and found the missing invoices
- **Names - algorithm did not find all the fuzzy matched names**
  - changed the algorithm to match all the names at the same time

Sample Final Table

GROWID1	GROWID2	MatchedOn	AddrOsa	NamesOsa	NumSplits	Dist	Agved
1	2	Addresses, lon-lat	3			0.05	TRUE
3	4	Invoices			108		FALSE

CONCLUSION

Importance:

- created a comprehensive list of account pairs that are matched
- used different algorithms to find pairs efficiently and effectively - can be used later to make data-driven decisions
- validates the agvend data and can provide updates
- versatile - ability to be used after the data has been updated continuously

Future Goals:

Marketing/Customer Analysis

- Use customer matches from final table to provide more accurate account data
- use final table for company insights into largest customers, customer groupings, and buying habits
- Run individually targeted marketing campaigns or give discounts to certain customers using customer buying habits

New Locations

- Grouped accounts for each customer would provide a better picture of where customers and farms are located.
- Use this information find new locations for the company and region to expand.

We would like to thank our TA's, Emily Weber and Harshini Musku, our Co-Alliance mentor, Tyler Netherly, our teams Data Scientist, Emily Hoeing, along with Tracy Soper, Jeff Griffeth, and all the Data Mine staff for their valuable input and guidance throughout the year.

References: https://stackoverflow.com/questions/51757806/partitioning-data-on-a-variable-to-speed-up-fuzzy-match-using-stringdist
https://rforjournalists.com/2020/12/15/how-to-access-open-street-map-in-r/