BUSINESS FORMATION STATISTICS (BFS) EFFICIENCIES

US Census Bureau | Business Formation Statistics | Economic Indicators Division

Kaustubh Harnoor, Data Engineer Fellow Fellow | *University of California, Santa Cruz, Computer Engineering*

Keywords:

Redundancies, Resource Reduction, Process Automation, Sustainable Algorithm Design

Summary:

I automated the manual process of adapting a machine learning algorithm for my team's objectives, streamlining the workflow and eliminating redundancies. These automations allowed us to processed 4-5 times more data, saving up to 20-25 hours each month. I also documented my revisions to the code to ensure seamless implementation and reuse in future years.

Business Formation Statistics (BFS) Efficiencies

Kaustubh Harnoor

Computer Engineering @ University of California, Santa Cruz

2024 CDF

Supervisor – Alyson Plumb, Taylor Spofford

Background: Organizational Structure

- Business Formation Statistics (BFS) is a team in the Economic Indicators Division (EID)
- BFS measures the number of new businesses to learn more about entrepreneurial activity in the US
- BFS utilizes a machine learning algorithm that predicts NAICS codes given a business description
  - GitHub - uscensusbureau/BEACON: Business Establishment Automated Classification of NAICS
  - Maintained by the Economic Statistical Methods Division (ESMD) team
  - This project required collaboration from ESMD and we are thankful for their support and feedback

```
NAICS codes with score > 0.01 for the business description 'restaurant':

NAICS     SCORE
-----     -----
492210    0.017431492207497463
561720    0.021507061856761 49
722511    0.4358003060020304
722513    0.4813180151544924
```

Background: BFS Annual Update

- On an annual basis, during their annual update BFS updates to the latest BEACON model

- On a weekly basis, BFS runs an adapted BEACON script to run the logic on business descriptions of EINs (Employer Identification Number) from the IRS (Internal Revenue Service)
  - EINs are assigned by the IRS to businesses in the US

- Sample Data:
  - EIN number: XXXXXXXXX
  - Business Description: “Coffee Shop”

Goals and Methodology

Initial Goals

- Make the BFS Annual Update process more efficient and manageable for future years
- Automate the time consuming and manual process of adapting BEACON

Secondary Goals

- Remove redundancies in our adaptation of BEACON

1. Efficiency

- Modularize BEACON code to simplify edits and started the 2024 Annual Update

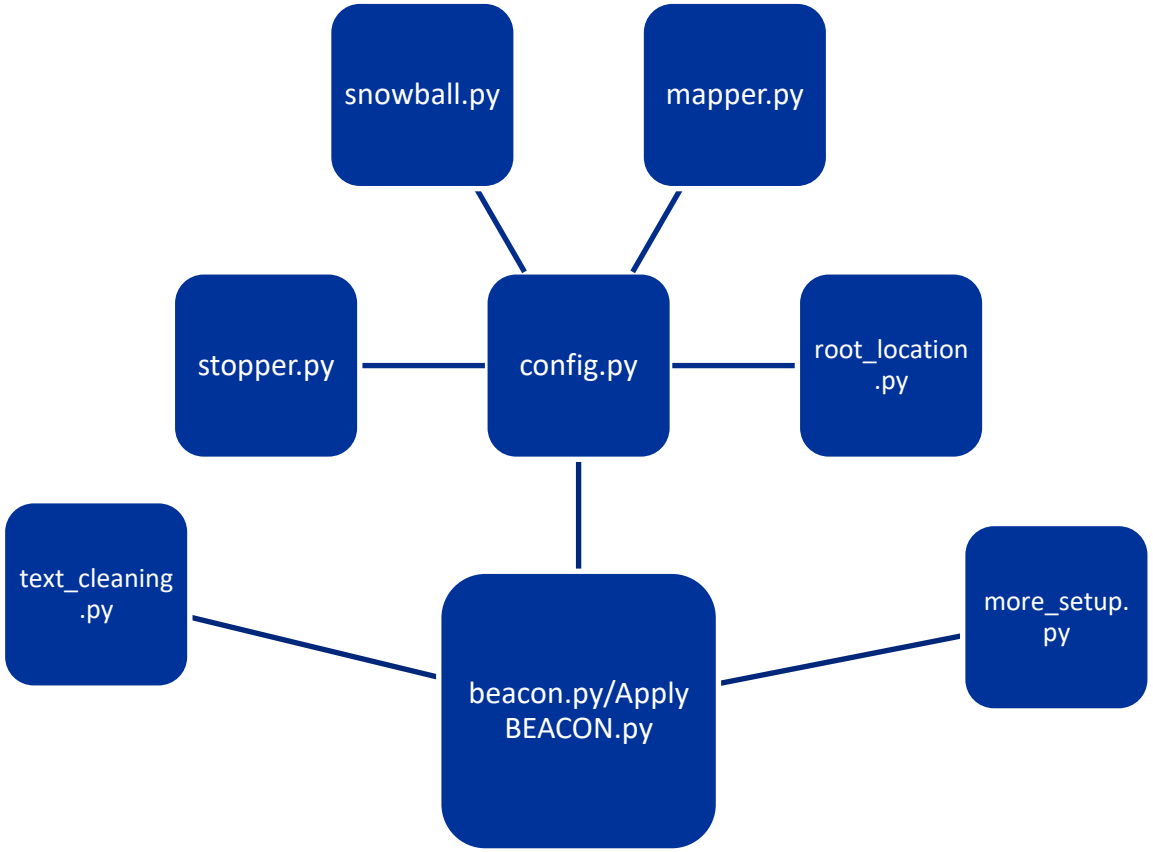
2. Automation

- Wrote instructions about the code we receive, our code, and on how to conduct the annual update

3. Redundancy

- Optimized the code to allow BFS to run it for a given time range (for example, a month) instead of weekly

Code Modularization



- Advantages
  - Reusability: Take BEACON's existing modules, keep them the exact same or have minor updates, and then add code for our purposes
  - Understandability: Becomes easier to manage our Annual Update if we treat the ESMD code as a "black box" that works and only change what's relevant to us
  - Efficiency: More time efficient because we aren't working with 20000+ lines of code

Cycle Optimization

- BFS can run Apply BEACON monthly instead of weekly

	Running Weekly	Running Monthly
Times Ran	52	12
Cycles (Weeks of IRS Data)	1	4-5
Run Time	2 mins 30 sec	2 mins 44 sec
Run Time per Year	130 minutes	33 minutes
Weekly BFS Schedule	56 - 70 hours per month	36.25 – 44.25 hours per month

19.75 – 25.75 hours per month saved

Documentation

Instructions for ESMD

- Modularizing the code
- Verifying outputs
- Ensuring the modularization does not affect the outcome

Annual Update Start

- Context about BEACON and Annual Update
- Modularization Details (files we receive from ESMD)
- Archiving Process

Code

- Purpose of Apply BEACON and step-by-step overview of the code
- Instructions for converting BEACON to Apply BEACON

THANK YOU!

Special thanks to **Alyson Plumb, Taylor Spofford, Eric Muro, and the ESMD team**!

If you have questions, I can be reached at
harnoorkaustubh@gmail.com