BUILDING A CENTRALIZED DATA PLATFORM FOR AGING-RELEVANT DATABASES
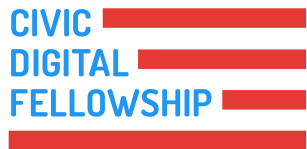
National Institute on Aging: Division of Aging Biology

Dr. Candace Kerr — PhD

CIVIC DIGITAL FELLOWSHIP

NIH

VINAY TUMMARAKOTA
Rice University
Computer Science

How can the Division of Aging Biology (DAB) identify the accessibility of DAB-funded databases to the public?

CIVIC
DIGITAL
FELLOWSHIP

RATIONALE

Suppose you are a researcher interested in receiving grant funding from the DAB...

MANUAL SEARCH

NAME	LINK	DESCRIPTION	CITATION
Primate Aging Database	https://primateda	A searchable database that contains basic health and husbandry data from primates across the lifespan and from numerous research institutions	
Human Mortality Database (HMD)	https://www.mort	Documents the longevity revolution of the modern era and to facilitate research into its causes and consequences	Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on [date])
Kannisto-Thatcher Database on Old Age Mortality	https://www.dem	Contains data on death counts and population counts classified by sex, age, year of birth, and calendar year for more than 30 countries; used to estimate death rates for older populations	Kannisto, Väinö Development of Oldest-Old Mortality, 1950-1990: Evidence from 28 Developed Countries. Odense University Press, Odense, 1994; ISBN: 87 7838 015 4
Study on Global Ageing and Adult Health (SAGE)	https://www.who	Longitudinal study collecting data on adults aged 50 years and older from nationally representative samples in China, Ghana, India, Mexico, Russian Federation and South Africa	
Osteoporotic Fractures in Men (MrOS)	https://na01.safe	Focuses primarily on risk factors and natural history of osteoporosis and fractures in community-dwelling men aged 65 and older	
AnAge	http://genomics.s	Contains longevity records for over 4000 organisms	Tacutu, R., Thornton, D., Johnson, E., Budovsky, A., Barardo, D., Craig, T., Diana, E., Lehmann, G., Toren, D., Wang, J., Fraifeld, V. E., de Magalhaes, J. P. (2018) "Human Ageing Genomic Resources: new and updated databases." Nucleic Acids Research 46(D1):D1083-D1090.
Comparative Cellular and Molecular Biology of Longevity Database	http://genomics.l	Contains a wide range of cellular and molecular trait data from 100 individual studies on vertebrate species that has been collected with the goal of identifying traits that are highly correlated with longevity	Stuart et al. A comparative cellular and molecular biology of longevity database. Age 2013 Oct;35(5):1937-47. doi: 10.1007/s11357-012-9458-y. Epub 2012 Jul 27.

ENSURING SUSTAINABILITY

- Manual search is **time-consuming**

- Manual search requires **additional** manpower

CIVIC
DIGITAL
FELLOWSHIP
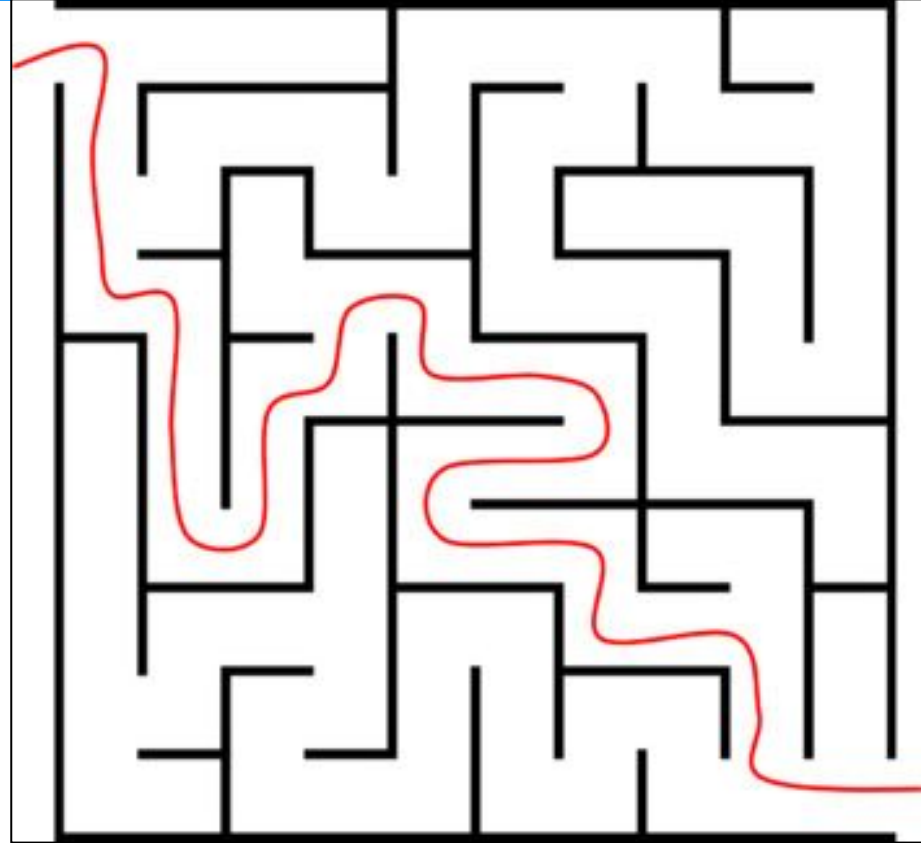
ENSURING SUSTAINABILITY

- Automate the collection of potential databases via **web scraping**

- Parsing the scraped results using **machine learning** to identify databases relevant to aging research

DEVELOPING A WEB SCRAPING TOOL

- Use specialized queries to search the **PubMed** Database

- Leveraged **BeautifulSoup** and **Selenium** to scrape the search results

(aging[Title/Abstract]) AND (database[Title])   ✕   **Search**

Advanced   Create alert   Create RSS   User Guide

Save   Email   Send to   Sorted by: Best match   Display options ●

174 results

☐ Reference **database** of hematological parameters for growing and **aging** rats.
1 Jacob Filho W, Lima CC, Paunksnis MRR, Silva AA, Perilhão MS, Caldeira D, de Souza RR.
Cite Aging Male. 2018 Jun;21(2):145-148. doi: 10.1080/13685538.2017.1350156. Epub 2017 Jul 11.
PMID: 28696817
Share CONCLUSION: The present study determined the normal values for absolute and relative hematological
parameters in Wistar rats from 2 to 24 months for male and female rats. The results can be used in
studies of effects of **aging**, feeding, and medications on growing and **agi** ...

☐ Modeling arterial pulse waves in healthy **aging**: a **database** for in silico
2 evaluation of hemodynamics and pulse wave indexes.
Cite Charlton PH, Mariscal Harana J, Vennin S, Li Y, Chowienczyk P, Alastruey J.
Am J Physiol Heart Circ Physiol. 2019 Nov 1;317(5):H1062-H1085. doi: 10.1152/ajpheart.00218.2019.
Share Epub 2019 Aug 23.
PMID: 31442381   Free PMC article.

APPLICATION OF A MACHINE LEARNING MODEL

Naive Bayes Classifier

APPLICATION OF A MACHINE LEARNING MODEL
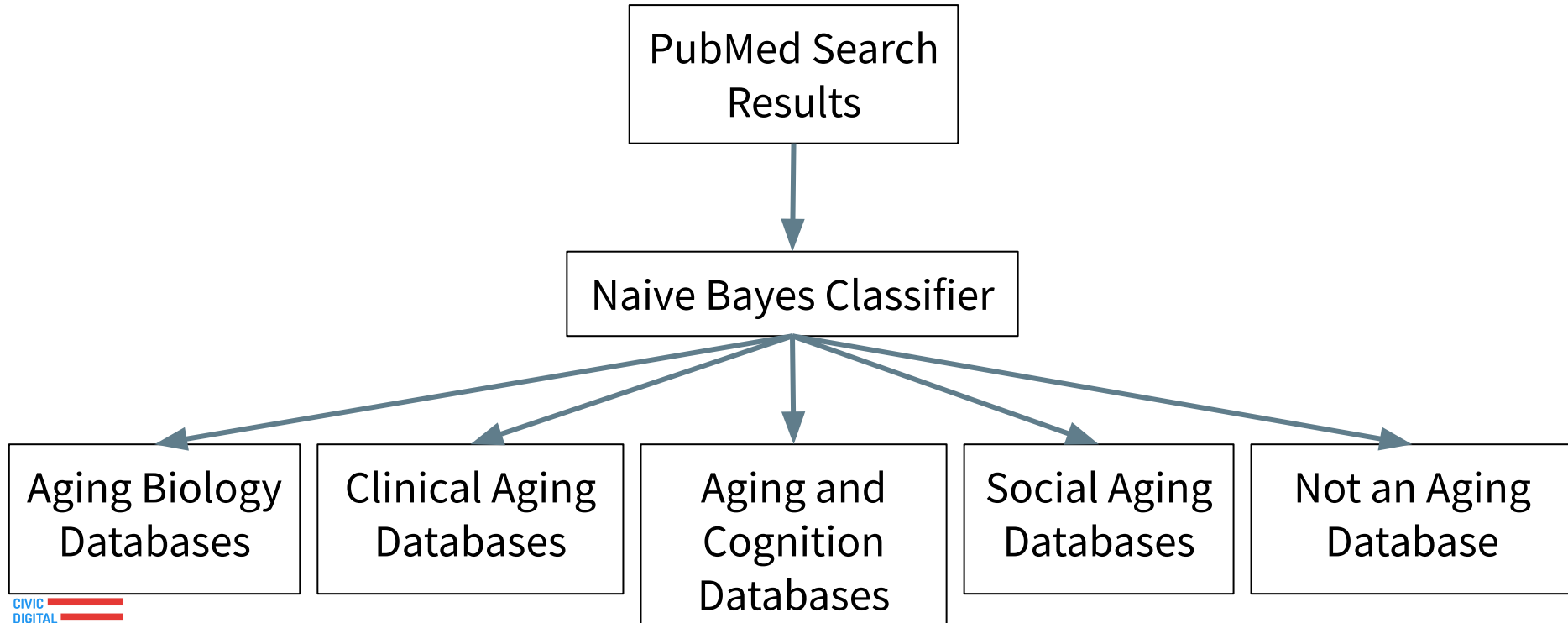
PERFORMANCE OF THE MODEL

Categorization of Databases (%)
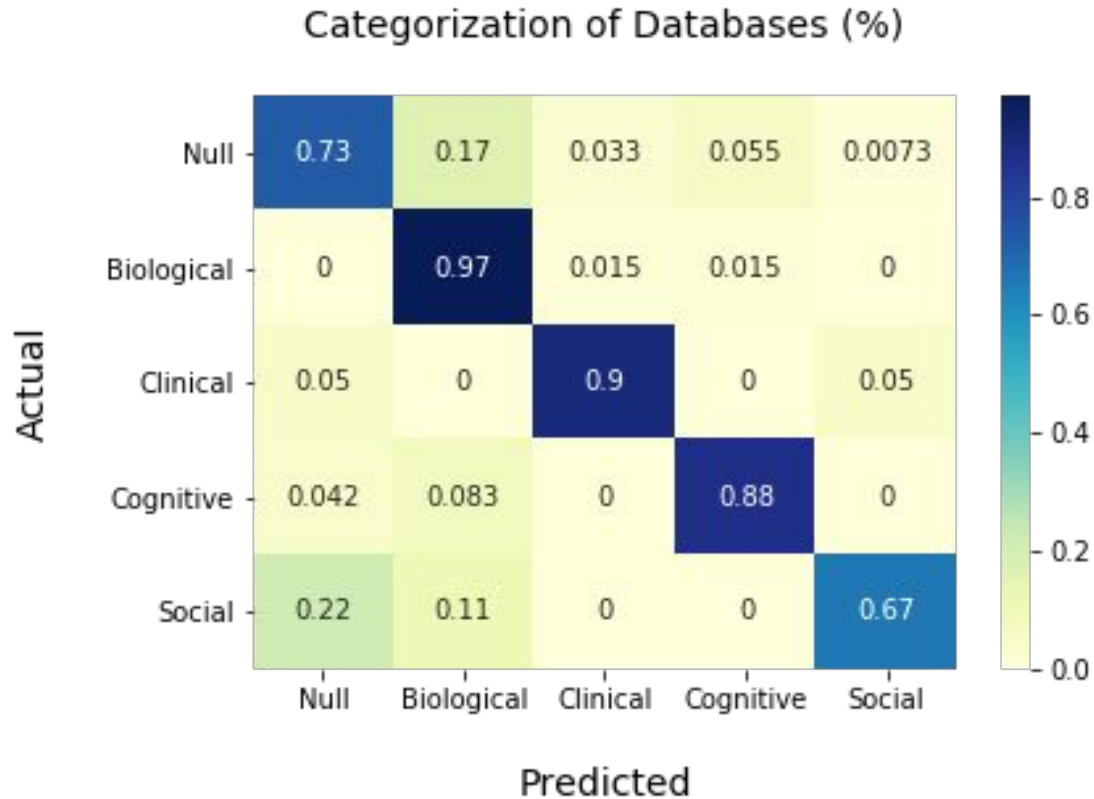
FUTURE DIRECTIONS

- Manually identified aging-relevant databases can be added to the NIA's Research Resources page

- Aging-relevant databases developed *in the future* can be quickly identified using web scraping & machine learning

ACKNOWLEDGEMENTS

- Dr. Candace Kerr

- Coding it Forward