Harvard Undergraduate Data Analytics Group

PREPARED FOR

NAACP Legal Defense and Educational Fund, Inc. (“LDF” or “Client”)

PREPARED DATE

Aug 29, 2023

ENGAGEMENT TIMEFRAME

September – December 2023

Harvard College Data Analytics Group (HDAG) is a non-profit student organization at Harvard dedicated to helping organizations make smarter and more data-driven decisions. We assist companies in achieving their strategic goals by translating their data into meaningful and actionable information. We aim to pair teams of well-trained, highly-motivated Harvard students with our partners, specifically focusing associates and analysts in industries where they have experience or interest, in order to produce the highest quality of work possible. From data collection to strategy implementation, we want to be there every step of the way to help organizations make data their new superpower.

We competitively recruit undergraduate students at Harvard with demonstrated competence, dedication, and problem-solving skills, many of whom have prior experience working in top management consulting or data science teams. All our team leaders have experience working in or leading data science teams at Fortune 500 companies, and our board of technical advisors include members of the Harvard faculty. Each team, composed of around seven to eight Harvard students, commits over 600 hours to a case over the course of a 10 week span.

We enjoy different challenges and work with a diverse set of organizations and problems. Our clients range from local businesses to Fortune 500 companies to international non-profits. Using our capabilities in visualization, machine learning, and predictive analytics, among others, we help organizations diagnose problems and identify strategies across their sales, marketing, financial or operational functions. Client confidentiality is our utmost priority.

Team Capabilities

1. Data Analytics Consulting: deriving valuable insights from data

- Case study 1 - Providing IT resource management analytics for a multinational Fortune 500 company in energy and automation: Through statistical analysis of over 100k anonymized employees, we identified help desk call volume and demographic trends to help inform executive decisions on employee satisfaction and IT resource allocation.
- Case study 2 - Providing data processing service for a Wall Street fintech company: Through scraping the Securities and Exchange Commission (SEC) website and extracting relevant data en masse, we created well-formatted databases to advance the client's core digital offerings.

2. Machine Learning Algorithms: training and deploying predictive models

- Case study 1 - Providing IT security service for a multinational Fortune 500 company in energy and automation: By building ML models, we enabled predictive analytics for the company's future spending on Indirect Procurements and introduced data integrity improvement design to the purchase request process.
- Case study 2 - Providing AI algorithm advancements for a leading sports analytics company: Using “Big 5” European club leagues’ pre-game and in-game data, we created models that predict win, loss, and draw probability and provided an evaluation of the accuracy and probability calibration of the models.

3. Business Intelligence Visualizations: creating interactive visual dashboards

- Case study: Providing visualization services for the World Health Organization Region for the Americas: We developed a web app to visualize models on COVID-19 outbreak to predict rate of transmission and epidemic curves; product delivered to WHO country offices in Latin America for projections of varying health intervention measures.

4. Whole-Set Solutions: providing comprehensive digitalization systems

- Case study: Creating an HR and user management system for an educational foundation in China: We developed a system from scratch to help the management team keep track of employee’s progress and KPI and to help employees better manage student feedback.

Proposal for LDF:

The goal of this project is to develop an application for monitoring and reporting election-related issues and suspected cases of misinformation and disinformation in real-time. Channels to monitor include, but are not limited to, election officials’ websites, news articles,

and Twitter. Broadly speaking, the application will be fed live data from predetermined sources, analyze and flag suspicious and anomalous content, and produce an alert/notification for each incident and or a recurring report summarizing incidents of the past time period to be reviewed for accuracy and follow-up actions manually.

The HDAG team will analyze and address the following specific items in their work.

- **Data Engineering and Collection:** The team will meet with Client to discuss current data available, data available upon request, and data to collect. The team will process, label, and build custom and client-provided datasets as necessary for the project. This data may include
    - Client's past datasets of misinformation sources to the degree that Client chooses to share them with the HDAG team
    - Reports regarding misinformation to the degree that Client is authorized to, and chooses to, share them with the HDAG team
    - Custom dataset built by scraping election-related news and non-private social media
    - Examples of true and false election information (e.g., election dates, specific policies, poll site locations).
- **Modeling:** The team will utilize machine learning and natural language processing techniques to build a model which predicts the truthfulness of given content. The team will explore a variety of methodologies to maximize accuracy and generalizability. Data will be vetted for accuracy and risks of bias based on race, gender, language, or other characteristics, including flagging content created by different groups at different rates and differential validity.
- **Application Engineering:** The team will build a software application which integrates the previously built model(s) and feeds it a private live stream of social media and news data. The team will explore methods for live alerts and the automatic generation of a report summarizing previous incidents utilizing Generative AI available to predefined and vetted users.

The team will thoroughly document their methodology and findings for all technical work throughout the duration of the engagement. A final presentation in slide format and a final writeup detailing the work, processes, motivations, and findings will be created and delivered to the client at the end of the engagement, including a plain language description of the model. The team will additionally provide all technical materials, including the code repository, engineered datasets, model-based findings, and final software product. Collectively, the foregoing shall be referred to as the "Services."

Rough Engagement Timeline

Dates	Week	Tentative Schedule
9.18-9.24	0	The HDAG Case Team Leader (CTL) will have a call with the respective Client liaison to better understand work expectations and align goals for this semester (in terms of research questions, final format of deliverables, etc.)
9.25-10.1	1	CTL will introduce the project and the work plan to the rest of the case team and start delegating tasks to each individual. (In each team we have data scientists who are proficient in Python, R, SQL and other analytical tools as well as business analysts who have experience working in industry).
10.2-10.8	2	Every member of each Client Case Team will follow the work plan, start both the data analytics, which includes every aspect of the data pipeline: data transferring, cleaning, exploration, modeling, visualization etc. Every week, the CTL will update the Client liaison on the progress that the case team has made over the past week. There is also a weekly meeting between the case team where each member will discuss their work with the others, and the CTL will delegate work for next week.
10.9-10.15	3	
10.16-10.22	4	
10.23-10.29	5	Wrap up the work for the first half of semester, and prepare for the midway presentation to Client which will include all progress made thus far.
10.30-11.5	6	Midway presentations with Client: the team will present their findings and recommendations from the first half of the semester to the Client team. The HDAG case team will follow up with any questions

		the Client team might have during or after the presentation.
11.6-11.12	7	After the midway presentations, the CTL will integrate comments or suggestions from the Client team to the work plan. The CTL will list out the remaining questions or technical tasks for the latter half of the semester and delegate them to each individual of the case team.
11.13-11.19	8	
11.20-11.26	9	
11.27-12.3	10	The case team will summarize their work for the entire semester and give a final presentation to Client. This will include both technical deliverables and the business presentation. The HDAG team will follow up with any questions the Client business team might have during or after the presentation.
12.4-12.17	Post- Project	HDAG team will follow up with Client on the implementation of suggestions and deployment of analytical tools. We will ask for feedback on their work for the Fall of 2023.

No Publicity. HDAG shall not use LDF's name, and any name of its subsidiaries or affiliates, or any adaptations of those names, for advertising, trade, or other commercial purposes without LDF's express prior written consent. HDAG and its contractors, employees and agents shall not hold themselves out as an employee, affiliate, or subsidiary of LDF at any time while performing the Services under this agreement. Any materials provided to HDAG by LDF pursuant to this agreement or in connection with HDAG's performance of the Services hereunder, bearing any LDF names, logos, styles, or trademarks may be used by HDAG only as necessary to perform the Services under this agreement. HDAG further understands and agrees that any violation or threatened violation of this section would materially and irreparably injure LDF and its business in a manner inadequately compensable in damages, and that therefore LDF may seek injunctive relief against the breach or threatened breach of HDAG's obligations herein in addition to any other legal remedies that may be available to it.

Independent Contractor. Each Party is separate and independent, and this agreement shall not be deemed to create a relationship of agency, employment, or partnership between or among them. Each party understands and agrees that this agreement establishes an independent contractor relationship and that the agents or employees of each respective party are not employees or agents of any other party.HDAG has all permits and licenses necessary to

perform the Services outlined in this agreement, as well as any insurance HDAG deems necessary. HDAG does not have any authority to enter into agreements or contracts on behalf of LDF and shall not represent that HDAG possesses any such authority.

HDAG is free to provide services to HDAG's own clients and to other parties during the term of this agreement. HDAG agrees not to perform services for HDAG'S own clients or other parties which may create a conflict of interest or interfere with HDAG's duties pursuant to this agreement.

HDAG will provide their own equipment for the Services rendered under this Agreement without reimbursement.

Termination for Convenience. This agreement may be terminated by either party at any time upon five (5) days' written notice.

Representations and Warranties. Both LDF and HDAG represent and warrant that each party has full power, authority and right to execute and deliver this agreement, has full power and authority to perform its obligations under this agreement, and has taken all necessary action to authorize the execution and delivery of this agreement. No other consents are necessary to enter into or perform this agreement.

Pricing

- Engagement Timeline: 12 weeks, September - December, 2023
- Semester Case Fee: Pro Bono

Signature: *Andrew Li*
Print Name: Andrew Li
Position: CCO, Harvard Data Analytics Group
Date: December 15, 2023

Signature:
Print Name:
Position:
Date: