Resources for Data Scientist, Machine Learning, and AI Scientist Roles for LatinxInAI/WomenInTech Community

Author: Isabel Metzger

Disclaimer: *These are my personal experiences within the interview process along with my code responses. The interviews may have changed but I wanted to provide the community with resources as much as possible!*

Outline Structure
- Company - Title
  - *Company and Title of Role Interviewed for*
- Description of Company & Role
  - *Brief description of Company & Role*
- Outcome
  - *Outcome of the interview, e.g., whether offer was provided, application rejected, or if withdrew from interview process*
- Overview of the Interview Process
  - *A brief overview of the interview process and any significant questions*
- Additional Thoughts
  - *Any additional thoughts of interview process*
- Additional Resources
  - *Description and name of file of additional resources, such as the pdf or code corresponding to the interview*

ExpressScripts - Machine Learning Engineer (Software Engineering Role)

Brief Description of Company & Role

Express Scripts is a pharmacy benefit management company.

Machine Learning Engineer with full Software Development Engineer status at the newly created Machine Learning Center of Excellence (MLCOE) within Express Script Inc's (ESI) Information Technology Department. Implemented machine-learning solutions at the production level and oversaw the code writing and modifying of complex software application programs of other team members, serving as a resource for team members with less experience.

Outcome

✅*Offer provided & I accepted*

Overview of Interview Process

I applied to this role by submitting it to the Grace Hopper portal. I was interviewed in-person while at the conference, Grace Hopper Women in Technology.
The steps required a whiteboard coding session and a 30-minute interview on my experience and interests.

Additional Thoughts

N/A

Additional Resources

N/A

Amazon Alexa AI - Applied Scientist

Brief Description of Company & Role

Express Scripts is a pharmacy benefit management company.

Machine Learning Engineer with full Software Development Engineer status at the newly created Machine Learning Center of Excellence (MLCOE) within Express Script Inc's (ESI) Information Technology Department. Implemented machine-learning solutions at the production level and oversaw the code writing and modifying of complex software application programs of other team members, serving as a resource for team members with less experience.

Outcome

✅*Offer provided & I declined*

Overview of Interview Process

I was contacted for this role sometime after finishing my thesis. I was provided several weeks to prepare before flying out to Seattle. AmazonAlex AI has a research center in Seattle and both Boston, but the offer was in Seattle.

The interview consisted of 2 full days and a preliminary recruiter assessment where you understand if your values align with Amazon's core values.

Coding was done on a whiteboard to assess the engineering skillset and understanding of logic. I was asked to present anything from prior work.

Additional Thoughts

N/A

Additional Resources

N/A

Datacamp Instructor

Brief Description of Company & Role

Datacamp is an online platform for learning machine learning and data science skills. Instructors are paid per lesson plans.

Outcome

✅ *Offer provided but had to reject due to company policy at ExpressScripts which did not allow me to get paid for teaching machine learning or NLP due to IP conflicts*

Overview of Interview Process

I was recruited by someone from Datacamp, I provided two lesson plans: one in python and one in R. Once the lesson plans were approved, I had to submit a 3-5 minute screencast covering a concept/topic in each outline and then complete slides based on the last lesson of Chapter 2, 3, or 4.

Additional Notes

This was an extremely fun and interesting experience. I wish that I was able to produce the lessons!

Additional Resources

- Accepted Curriculums
  - Deciphering Customer Feedback with NLP in Python - *outline_datacamp_Deciphering_customer_feedback_with_NLP_in_Python.pdf*
  - Neural Networks in R - *outline_datacamp_Neural_Networks_R.pdf*

Roivant - AI Scientist

Brief Description of Company & Role

Joined the Computational Research Team, creators of drugOME, as the company's first AI Scientist. My position as AI Scientist with experience in Natural Language Understanding applied to information extraction to help build and optimize the unstructured data ingestion pipeline that underlies our pharmaceutical analysis engine. This pipeline enables the aggregation of information about drugs across various dimensions. The data sources are diverse in origin and massive in scope, requiring classification, ranking, information extraction and linkage to enable drug identification and monitoring of the scientific progress made in disease treatment.

Outcome

✅ *Offered & Accepted*

Overview of Interview Process

1. 45-minute phone screen with the director of NLP and AI
   a. *This was a time for the director to discuss the NLP team and hopes for the role as well as learn about my experiences*
2. Thought Exercise - take-home prompt with 5 day-7day timeline (more details in *RoivantThoughtExercise.pdf*)

High level description of the exercise:
1. Build dictionaries of MoAs/diseases synonyms from the MeSH raw data files (coding exercise).
2. Build a knowledge base from the co-occurrence of diseases/MoAs tagged in PubMed abstracts to establish that a relationship may exist (coding exercise).
3. Propose a strategy leveraging NLP tools identifying MoAs/diseases and their relationships to build our knowledge base (non-coding exercise)

3. Interview In person
   a. 1hr presentation of Thought Exercise to team consisting of:
      VP of Computational Research, Director of Computational Research, 1 Data Scientist, 3 Data Engineers
   b. 30 minute 1:1 interview with Data Scientist
   c. 30 min 1:1 interview with VP
   d. 30 min 1:1 interview with Data Engineer
   e. 30 min 1:1 interview with Director
   f. Wrap up with the HR and Talent Acquisition Manager

All 1:1's were very casual and provided a chance to learn about each member of the team and their broader role within the company.

Additional Thoughts

This was an incredibly fun interview and I really enjoyed presenting and meeting everyone 1:1. I was extremely impressed with the amount of free-text data they had been mining over the years. One thing that the interviewers said they really enjoyed was my interactive visualization for topic modeling.

Additional Resources

- Description of Role - *RoivantThoughtExerciseAIScientist.pdf*
- Thought Exercise Description - *RoivantThoughtExercise.pdf*
- Roivant Code for Thought Exercise - *RoivantCodeThoughtExerciseIsabelMetzger.pdf*
- Part 3 Write-Up - *RoivantInterviewAIScientistPart3-IsabelMetzger.pdf*

Covera Health - AI Scientist, Deep Learning for Natural Language Processing

Brief Description of Company & Role

Startup based in NYC focused on radiology and EHR. AI Scientist to work on their proprietary platform.

Outcome

❌*Application Rejected - Made it past the 1 and a half-hour presentation of previous work, feedback provided was that my presentation missed some of the things they were looking for*

Overview of Interview Process

1. 45-minute phone screen with Chief AI Officer
2. Paired Coding Interview with Sr AI Scientist (NLP) and AI Scientist (NLP)
    a. Prompt provided the night before via GitHub
3. 1.5 hour-long presentation of previous work (see presentation description below)

The presentation can be either an overview of projects you have worked on or one project you were involved in stating the problem, the objective, and your solution. In both cases, you should emphasize your contribution. It could be your thesis but it could also be anything on whatever you've published so far (if applicable.) We do ask that at least one of the projects be related to NLP and should follow the format of a conference talk. For instance, a suggested outline of the presentation is:
INTRODUCTION: Why was this study/research performed? What unsolved problem are you addressing?
METHODS: How did you study this problem?
RESULTS: Report the data, analyses, and/or outcomes

DISCUSSION: How do you interpret the results?
CONCLUSION: What is the relevance to clinical practice or future research?

Additional Notes

N/A

Additional Resources

- Job Description - *CoveraHealthAIScientistJobDescription.pdf*
- Paired Coding Assessment Prompt -
  CoveraHealthAIScientistPromptPairedCodingInterview.pdf

Flatiron Health - Senior Data Scientist, Real World Evidence

Brief Description of Company & Role

Data Insights is a diverse and talented team with strong expertise in discovering, prototyping and launching new data products. As a Senior Data Scientist on this team, you will define data strategy, drive technical execution, and help us create the next generation of our real-world evidence data products. You will be embedded on a newly formed team within our Real World Evidence business line. This team is focused on developing a new set of large-scale, ML-enabled research data products. You will be tasked with creating high quality clinical research variables from structured and unstructured data sources using creative approaches to machine learning and natural language processing. You'll work cross-functionally with software engineers to productionize these new variables, and with quantitative scientists and oncologists to validate that they can be used to generate sound scientific insights.

Outcome

❌*Application Rejected - Provided an answer over the phone saying that team did really like me*

Overview of Interview Process

1. Initial screening is a take-home code assignment to be completed within 7 days
2. The second interview is a live coding session
3. Full-day interview consisting of:
   a. 1 hour with two data scientists (Data Science-based interview)
   b. 1-hour live coding interview (a second live coding interview)
   c. Collaborative interview for picking and choosing a data scientist project
   d. Interview with Upper Management

Additional Notes

This was the only company I had to sign an NDA for the full-day interview.

Additional Resources

- Job Description - *FlatironSeniorDataScientistRWEJobDescription.pdf*
- Code Submission 1 - *FlatironExerciseInitialAssessmentIsabelMetzger.pdf*

NewtonX - Data Scientist

Description of Company & Role

NewtonX matches the world's toughest questions with the only minds who know the answers. We find the right subject matter expertise from an open network of 1.1 billion professionals. To do this, we use the most sophisticated search engine in the research industry – the NewtonX Graph. It identifies the exact audience for our clients' business questions. We ensure every professional is 100% verified, so you can ground your work in true expertise and make strides with confidence.

As the world's leading B2B research company, we field large-scale quantitative surveys, facilitate qualitative interviews, engage in long-term consultations, and create customized research plans. NewtonX partners with the world's top consultancies, marketers, and technology companies. Together with our clients, we're ushering in a new standard of truth in B2B research. To learn more about us, head to newtonx.com.

We are seeking Data Scientists to join our team in shaping the direction of data science across a wide variety of problems, including our proprietary knowledge graph, search functionalities, and more. Our team is smart, visionary, supportive and, above all, excited by what we do! We are looking for strong, inquisitive data scientists with clever solutions, with a desire to join a small but fast-growing team (<50) to get in early and help shape the direction of our technology.

Outcome

❌Application rejected, feedback provided from the recruiter below:

"YP at NewtonX enjoyed speaking with you and thinks you are very smart and passionate.

He thinks you are more interested in a research-led position where you can make a social impact and this isn't the main priority for NewtonX. They interviewed someone at the final stage this week and have made an offer."

Overview of Interview Process

The first initial interview was a 45-minute video interview with the Chief Technology Officer about the role and my experience.

Additional Notes

This is where I believe my focus on explaining the methodology of the projects rather than emphasizing the main outcomes may have hurt my application.

Additional Resources

- Job Description - *NewtonXData ScientistJobDescription.docx*

DataRobot - Machine Learning Engineer

Description of Company & Role

Machine learning engineer to focus on automating data science and ml workflows.

Outcome

⚪Withdrew from the interview process because I had another offer

Overview of Interview Process

30-minute phone screen and then a take-home coding assessment.

Additional Resources

- Take-Home Coding Assignment Instructions - *DataRobot-Machine_Learning_Engineer_-_Take_Home_Test.pdf*
- Preliminary Code for Assessment - https://github.com/izzykayu/DefaultLoanPrediction

---

EVERSANA - Data Scientist reporting to the Chief Analytics Officer

Description of Company & Role

EVERSANA provides B2B services and consumer services using data analytics analyzing and understanding the life and health sciences.

Embedded in the Data Analytics and Business Unit as a data science professional reporting directly to the Chief Data and Analytics Officer. Design, implement and deploy advanced machine learning / artificial intelligence algorithms across multiple domains such as healthcare, media, retail, and manufacturing.

Outcome

✅ *Offer Provided & I accepted*

The first interview was a 1 hr with the Chief Analytics Officer and was a review of experiences and the vision for the role and team, the second interview was a 1hr technical interview doing a case study together with a Data Science Manager, and the third interview was also a 1-hr technical case study with the Director of Data Science, with much more focus on the statistical evaluation process. The case studies were healthcare-related: e.g., identifying potential patients of rare diseases

Additional Resources

N/A

HealthFirst - Data Scientist

Brief Description of Company & Role

Healthfirst is a non-profit company committed to providing patients links to services in healthcare.

Outcome

○ Withdrew from the interview process because I had another offer

Overview of Interview Process

1. 1-hour interview with Lead Data Scientist and interim head of Data Science
2. 1-hour interview with Director of Epid
3. 30-minute interview with Data Scientist
4. Take-Home Assessment with a week to complete
5. 1-hour presentation/panel session (Chief Analytics Officer and other Data Science team members present)

Additional Thoughts

Questions included "Where do you see yourself in 5 years?" and other types of behavioral and personal goals questions.

Additional Resources

- Description of Role - *HealthfirstDataScientistJobDescription.pdf*
- Assessment Exercise Description - *HealthFirstDataScientistCandidateTakeHomeActivity.pdf*

US Government - Data Scientist

Brief Description of Company & Role

Outcome

✅ *Certificate provided and open until 2023 to interview with different agencies*

Overview of Interview Process

Resume review: Subject Matter Experts will review the first three (3) pages of your resume to determine your technical qualifications for this position based on the required competencies.

Work Sample Assessment: The work sample assessment will include analysis of multiple public datasets totaling 150 MB or less in size. Your response to the work sample assessment will be used to measure your depth of knowledge in relation to the required competencies for the position.

Qualifying phone interviews: At this phase, qualified candidates will have at least one phone interview to further assess whether their experience meets the required competencies. Qualified candidates may meet with more than one Subject Matter Expert for that determination to be made.

Additional Notes

The coding 1 hr Hirevue interview with Subject Matter Expert consists of 3 questions. This was a difficult interview for me because without seeing the interviewer's face, it was more difficult for me to have the cues for if I answered something in an expected way.

Additionally, the coding tasks (sample assessments) definitely took more time than the couple of hours they had expected, that being said I thought the process was really cool: the data was real, messy, full of missing values, etc., the government database that was relevant to real world analysis. The hiring process was very cool because you have up to 2 years to interview after being provided the certificate to be considered for Data Scientist roles.

Agency Specific Interviews

Interviewing with several agencies almost felt like "speed dating". There were a few agencies that offered a quick superior qualifications package but I held out for DHS. DHS was the most exciting--it was really great to meet the team I would be working with. I really felt that out of all the agencies and teams, this could be a team where I could distinguish myself and stay for the rest of my career. Prior work using NLP and ML to mine through reports to help investment analysts at a biotech startup seems to be not related too much, but still in the realm of finance.

Questions that stood out to me: a question regarding a time where I disagreed with another colleague on an approach and how did I communicate and work through that.

Outcome

✅ Multiple agencies offer including State Department and DHS (GS-13 Step 08), others GS-13 Step 01

Additional Resources

- Description of Role -USAGov*JobDescriptionDataScientist.pdf*
- Assessment Exercise Description - *USAGOVDataScientistPrompt.pdf*
- Submission of Data Scientist Work - *IsabelMetzgerUSAGOVDataScientistSubmissionFeb1st2020.pdf*
- Code on EDA of Fatality Report - *USAGovDataScientist_EDA_Analysis_Fatality_Detroit_NYC.pdf*

Department of State - Data Scientist

Brief Description of Company & Role

Outcome

Overview of Interview Process

Resume review: Subject Matter Experts will review the first three (3) pages of your resume to determine your technical qualifications for this position based on the required competencies.

Work Sample Assessment: The work sample assessment will include analysis of multiple public datasets totaling 150 MB or less in size. Your response to the work sample assessment will be used to measure your depth of knowledge in relation to the required competencies for the position.

Additional Notes

Outcome

- Description of Role -USAGov*JobDescriptionDataScientistSTATE.pdf*
- Assessment Exercise Description - *USAGOVDataScientistPromptSTATE.pdf*
- Submission of Data Scientist Work - *IsabelMetzgerUSAGOVDataScientistSubmissionFeb1st2020.pdf*
- Code on EDA of Fatality Report - *USAGovDataScientist_EDA_Analysis_Fatality_Detroit_NYC.pdf*

DCSA - Data Scientist

Brief Description of Company & Role

Outcome

Overview of Interview Process

Resume review: Subject Matter Experts will review the first three (3) pages of your resume to determine your technical qualifications for this position based on the required competencies.

Work Sample Assessment: The work sample assessment will include analysis of multiple public datasets totaling 150 MB or less in size. Your response to the work sample assessment will be used to measure your depth of knowledge in relation to the required competencies for the position.

Additional Notes

Outcome

Additional Resources

- Description of Role -USAGov*JobDescriptionDataScientistSTATE.pdf*
- Assessment Exercise Description - *USAGOVDataScientistPromptSTATE.pdf*
- Submission of Data Scientist Work - *IsabelMetzgerUSAGOVDataScientistSubmissionFeb1st2020.pdf*

- Code on EDA of Fatality Report - *USAGovDataScientist_EDA_Analysis_Fatality_Detroit_NYC.pdf*