**Compass Dashboard Project**

**Q1. Tell me about you current Project.**

Ans: The Compass Dashboard Project is to working with the Employee’s details (When Employee join the organization their Personal as well Professional details as been noted in the Report) and **Goal** is looking into the output of Employee productivity in the reskilling, IJP, Project, Transforming Employee’s from comparable > Efficient > Excellent in the Skills and Skill tag, Rating purpose.

**Features of Compass Dashboard**

* Unique hits count for the compass profile.
* Compass profile completion percentage.
* Opportunities(IJP’s and Project Demands)
* Careers in Digital tags and Aspired career path.
* Mentoring, Mentees and relationship established.
* Learning stacks progress.
* Training program Quarterly update & Certificate completion.

**Functionalities in the Compass Dashboard.**

Compass Dashboard includes 7 Subsets.

* Cube data – Capture the Compass profile hits and accessibility.
* Compass Profile completion includes skills update.
* IJP and Project Demand – Eligible check and project change
* Mentoring – Eligible for mentoring and mentees looked for mentor.

**Deployment/Installation**

* The application can be easily installed as a web-based API on any cloud platform. This installation is similar to a Zoom app

**Q2. What was the size of the data?**

**Ans**: 2.5 L rows of CSV report is used to category the data.

**Q3. What was the data type?**

**Ans:** The data type is CSV file format.

**Q4. What was the team size and distribution?**

**Answer:** The team consisted of:

* 1 Product Manager
* 1 Solution Architect
* 1 Lead,
* 2 Dev-Ops engineers
* 2 QA engineers
* 2 UI developers, and
* 3 Data Scientists

**Q5.What Hadoop distribution were you using?**

**Answer:** we have implemented HDFS and MapReduce for a well-known learning algorithm—decision tree in a scalable fashion to large input problem size. Computational performance with node count and problem size is evaluated.

**Q6.What is the version of distribution?**

**Answer:** CDH – 5.8.0

**Q7.What was the size of the cluster?**

**Answer:** The cluster(production setup) consisted of 15 servers with

* Intel i7 processors
* 56 GB of RAM
* 150 GB of Secondary storage each

**Q8. How many nodes were there in all the Dev, UAT, and Prod environments?**

**Answer:** The development code part run in individual developer’s workstation. Duplicate copy of development code will be run in the cloud platform for testing and after testing done that model need to be used for further classification. Those models later will be executed in project deployment of Cloud UAT and Prod Environment.

In total, we had:

* 5 nodes in the dev environment,
* 5 nodes in UAT, and
* 10 nodes in production.

**Q9.How were you creating and maintaining the logs?**

**Ans:** The login through Company’s mail Id as username and password.

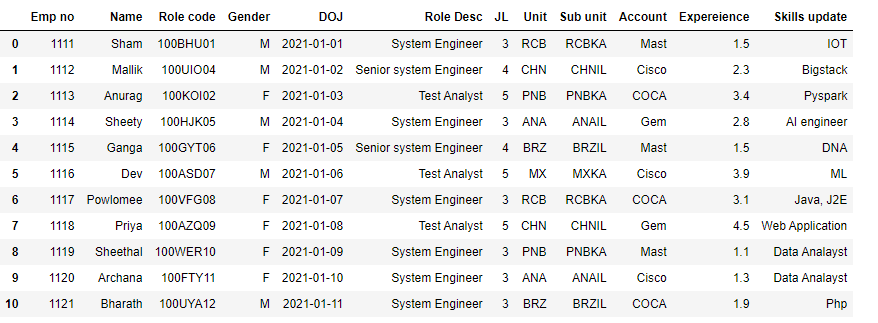
If anyone tried to do login with other than company’s credential throws Error and block it after 3 times entered

**Q10.**What techniques were you using for data pre-processing for various data science use cases and visualization?

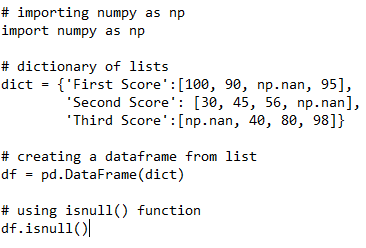
**Machine Learning**

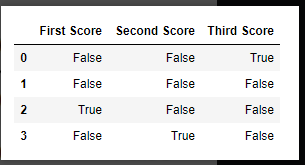
It is a data mining technique that transforms raw data into an understandable format. Raw data(real world data) is always incomplete and that data cannot be sent through a model.

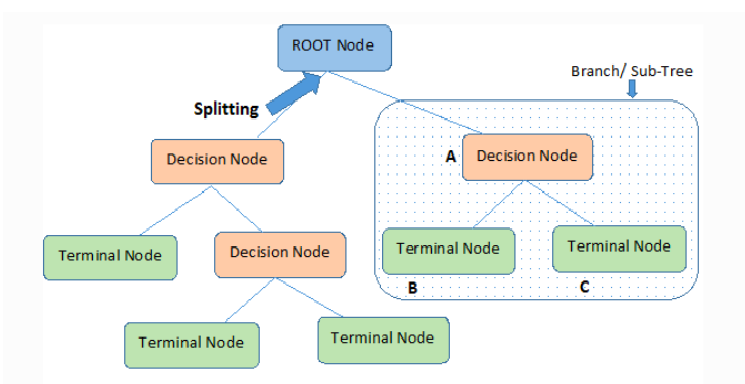
* Import Data, Read Data, and Checking for missing values, Checking for categorical data.
* Standardize the data, PCA transformation, Data splitting.



* Finding missing values







**Q11. How were you maintaining the failure cases?**

**Ans:** Compass profile loggin model only accepts only company’s mail ID other mail ID it will show authentication error. It will be showing Wrong user ID after 3 time It will get block.

**Q12.What kind of automation have you done for data processing?**

**Answer:** Decision trees use multiple algorithms to decide to split a node into two or more sub-nodes. The creation of sub-nodes increases the homogeneity of resultant sub-nodes. In other words, we can say that the purity of the node increases with respect to the target variable. The decision tree splits the nodes on all available variables and then selects the split which results in most homogeneous sub-nodes.

**Q13.Have you used any scheduler?**

**Answer**: yes after 30 days models are retrained

**Q14.How are you monitoring your job?**

**Answer**: we have kept on observation for certain period in mean while if any issue occurred mail will get trigger for us in the notification.

**Q15. What were your roles and responsibilities in the project?**

**Answer:** Collection of datasets undergo as data batches for testing model , than deployment done in the cloud platform and to be monitored the model for issue occurred and provides a QA check and support before final deployment and share with the client.

**Q16.What was your day to day task?**

**Answer:** My daily work routine is checking the mail, working on the queries, meeting with team for any modification in projects or monitoring the project and supporting the design for the project .

**Q17.In which area you have contributed the most?**

**Answer:** I worked much on the Data validation , checking for the suitable algorithm for use cases. Spending most of the part on debugging , implementation of deployment of project.

**Q18.In which technology you are most comfortable?**

**Answer:** I have been a part of AI, ML DL and NLP projects but I love to work on ML Projects.

**Q19.How you rate yourself in big data technology?**

**Answer:** As I got opportunity to work with small models aware of working and ability to modify the data if got any time to work on that I will take that chance.

**Q20. In how many projects you have already worked?**

**Answer:** I have worked in small projects as part of data transformation, data validation, deployment and model load for specific cluster, data preprocessing. And we have used Decision tree Machine learning for the Compass Dashboard project.

**Q21. How were you doing deployment?**

**Answer:** The steps I followed to deploy Compass Dashboard. Our stakeholder wanted to deploy the dashboard on the cloud based. In the Dashboard we have many models but here we showing one model deployment i.e.., Compass profile. Before deployment we prepared models and check with API for proper output for the decision making.

**Q22.What kind of challenges have you faced during the project?**

**Answer:** The challenge I faced was matching the reports which can give need able output and data trimming converting to good dataset available for my project and selecting algorithm and when uploaded model to API error all these were part of my work.

**Q23.What will be your expectations?**

**Answer:** I have worked small projects also part of big projects as per my learning & experience, career growth is also one of the part. I want work on new project to explore much more learning idea involves me.

**Q24. What is your future objective?**

**Answer:** Increase professional knowledge and training as well Increase earnings.

Developing growth mindset more good habit of reading. Have interest to work in End to End project.

**25. Why are you leaving your current organization?**

**Answer:** Looking for new challenges as well career growth. Wanted to grow in the similar way of the market growth. Current organization is best one to work but similar kind of work and project I used to do so need change.

**Q26. How did you do Data validation?**

**Answer:** Data validation is when a program checks the data to make sure it meets some rules and restrictions. Each model need to do login with credentials it will show authentication error if we use outside email id while doing login.

**Q27. How did you do Data enrichment?**

**Answer:** Apart from Data enrichment , we done extraneous data can introduce distractions and confusion, so its counter productive to enrich existing data with additional information that isn’t necessary for our purpose. Enhancing collected data with relevant context obtained from additional sources.

**Q28. How would you rate yourself in machine learning?**

**Answer:** 7

**Q29. How would you rate your self in distributed computation?**

**Answer:** 7

**Q30. What are the areas of machine learning algorithms that you already have explored?**

**Ans:** I have Explore on multiple projects Machine learning, Natural Language Process, Deep Learning, K-Means, Wanted explore more in other kind of projects.