



AI/ML Solutions in Product Management

Capstone Project

Submitted by: ACHAL KAGWAD

COHORT-C13

Date: 15th March 2023

achal.kagwad@gmail.com | 9108302174

After completing all the parts of the project, you will have to submit your responses in the following manner.

- Submit the following parts of the project in this **“Presentation”** file

Parts	Topic	Weightage
Part 1	Understanding the Market	5%
Part 2	Competitive Analysis	5%
Part 3	Business Objectives	5%
Part 4	User Research	5%
Part 5	Product Artifacts	10%
Part 6	Reimagine Your Product	10%
Part 7	Business Benefit Estimation	10%
Part 8	Data Setup	20%
Part 9	Selection and Measure of Model	20%
Part 10	MVP	10%

Part 1: Understanding the Market

- **TAM(Total Available Market):** Total Population of India is 1.4 Billion ie 140 Crores. Considering urban and rural population to be in the ration of 30: 70, Approximate urban population is 42 Cr and the rural population is 98 Cr.
- **SAM (Serviceable Addressable Market):** Daily Hunt wants to serve customers with access to fast internet who can read English and Local Languages. Assuming 90% of Urban population have access to fast internet and 70% of rural population have access to fast internet we have $0.9 \times 42 \text{ Cr} = 37.8$ and $0.7 \times 98 = 68.6 \text{ Cr}$. Total of $37.8 + 68.6 = 106.4 \text{ Cr}$
- **TARGET AUDIENCE:** As per our business model, we filter on age group of 20 to 45 who have a smart phone. Assuming 70% of Urban population who have access to fast internet and in the desired age group and 80% of Rural Population with similar filters: $(0.7 \times 37.8 + 0.8 \times 68.6) = 26.46 + 54.88 = 81.34 \text{ Cr}$.

- **Threat of New Entry: Medium Force**

1. Capital Requirements: Daily Hunt has raised a total of 1.7 Billion till date. Thus new entrant has to come with deep pockets. Entering this market requires long term(in years) planning, It takes years to have regular customer base.
2. Technology and Infrastructure costs keep scaling with new users, Skilled Tech Talent demands high CTC
3. The four phases to hook a customer: Trigger, Action, Variable Reward and Customers Investment into app takes time and planning.
4. New Entrants like way2news, News Bytes etc need to build a network of content partners which takes cost, time and effort.
5. However having said this, its not entirely impossible for new entrants to capture niche and scale vertically or horizontally in future.

- **Threat of Substitutes: Strong Force**

1. We are not only competing with news aggregators but also for the attention span of users in all other areas which demand their attention such as entertainment, audio and video. We have low switching cost.
2. Hyperlocal short form news aggregators are increasing with local Indian Languages. Hence loyalty will decrease further.
3. Local Hard Copy News Papers also demand attention of users and users can switch to them for quick browsing of headline news.

- **Bargaining Power of Buyers: Strong Force**

1. Cost of Switching is low since most news aggregator apps are free to use.
2. Availability of Substitutes.
3. Driving DAU and MAU is a challenge

- **Bargaining Power of Suppliers: Moderate Force**

1. Until user is hooked to our platform we have low supplier bargaining power.
2. The Product Offered by other Creators are not very niche and unique.
3. With 100K content partners and creators who publish 250K news and content pieces every day in 14 languages and a massive user base for feedback. With proper use of data, AIML technology to personalize content we can drive this bargaining power to a Strong Force.

Porter's Five Forces Template

upGrad

Threat of Entry
"Medium Force"

Bargaining Power of Buyers
"Strong Force"

**Competitive
Rivalry**
"Strong Force"

Bargaining Power of Suppliers
"Moderate"

Threat of Substitutes
"Strong Force"

Part 2: Competitive Analysis

Competitor Analysis- Strengths and Weaknesses

CRITERIA	FLIPBOARD	GOOGLE NEWS
Strengths	<ul style="list-style-type: none">• Clear attractive user interface – Visually appealing and stunning UI. Instead of scrolling you will flip pages. Thus called Flipboard.• High Retention Rates• Recommendation systems in app.	<ul style="list-style-type: none">• Power of ecosystem and collaboration with other google apps. Account Creation and integration is seamless.• Fast Expansion: Already 1Bn+ total downloads, thus more reviews and stronger marketplace.• Brand Equity Presence
Weaknesses	<ul style="list-style-type: none">• Low Brand Recognition• User Journey with Account Creation not intuitive• Search Engine not as efficient as google news	<p>Visuals are not great with minimalistic design.</p> <p>Competition from local news discovery apps</p>

CRITERIA	FLIPBOARD	GOOGLE NEWS
App Downloads	500MN+ on Android play store.	1Bn+ on Android Play Store
App Ratings	3.9 stars on play store	4.2 stars of play store.
UX of App	Design Of App: Intuitive, Clear Root Icons at the bottom indicating their functionality. Has various features such as user personalization with recommendations.	Minimalistic Design. However the home page scrolls down indefinitely where there can be definite end. Overall the app UX is neat.

CRITERIA	FLIPBOARD	GOOGLE NEWS
Browsing User Stories	Top 10 for today displays top 10 news that may need user attention. Flipboard also uses recommendation techniques with its own user data.	Google uses user data and presents them in “For You” Tab as recommendations. Personalisation increases retention.
Customization	Confined User Space, No scope of user mind to deviate to other apps and demands user attention.	Just feels like other Google App and high tendency of user to switch to other apps during the reading process.
Curating and Sharing	Design Of App: Intuitive, Clear Root Icons at the bottom indicating their functionality. Features such as category wise segmenting and curating. History of Curated Articles.	Minimalistic Design. Sharing of News not very intuitive. Curating and storing of articles may be improved. Overall the app UX is neat.

CRITERIA	FLIPBOARD	GOOGLE NEWS
Predict Future Initiatives with Reasoning	Video content: With the growing popularity of video content, Flipboard may invest more resources in incorporating videos into its platform. This could include partnering with video content providers, creating its own video content, or introducing new features to make it easier for users to find and share video content.	Personalized content: This could include features such as customized news feeds, personalized alerts, and improved recommendations. Improved news curation: using machine learning algorithms to identify and surface high-quality sources, as well as tools to help combat fake news and misinformation.
	Augmented Reality: or example, it could use AR to create interactive experiences that blend digital content with the real world, or to provide users with more immersive ways to explore and consume content. Community features: improving its social networking features by introducing new ways for users to connect and share content with each other, such as group discussions or	Integration with other Google services: Google News could become more integrated with other Google services, such as Google Assistant, Google Maps, and Google Search- receiving news updates while commuting or accessing news related to a particular location. AR/VR news experiences: to provide real-time contextual information about

Part 3: Business Objectives

1. Build Multi Language Data Analytics to generate insights for business and content generators.
 2. Increase hyper local content consumption driven by content in Indian Languages.
 3. Identify Popular categories for Indian Language Internet Users and build content basis that.
 4. Voice Recognition and language agnostic internet: Advanced voice recognition and translation technology could help Indian language internet users.
- **Problem Statement:** “How to Leverage AI/ML solve to solve the problem of decreasing user engagement and increasing time spent on news aggregator platforms”

Part 4: User Research

OBJECTIVES

- Identify and create audience segments of users based on various factors/features such as demographic profile, user reading habits, preferences, impression caps, duration engaged etc to increase user engagement, reduce churn and drive revenue.

HYPOTHESES

- Users prefer to read online over newspapers
- Users focus more on local news over world news
- Weather report is checked by every user who uses the app
- Category most read is sport news over movie news
- Users would like to have pop ups for breaking emergent news
- Users prefer to have simple UI rather than flashy bright UI
- Users prefer flipping page rather than scrolling down
- Users prefer to read news in batches whenever they are free during the day
- Based on the age and career phase users in mid 30S prefer immersive news for wealth management and just headlines or short news for sports and Bollywood.

Interview Questions:

- Do you use Daily Hunt App to get news?
- How often do you read news on the app on a daily basis?
- What is the most watched section in the app?
- What kind of news you usually read, Sports or Bollywood or anything else?
- How do you want emergency news to be communicated?
- Do you prefer to read the news in English Language or your regional language?
- What UI style do you prefer, flipping pages or scrolling down?
- Do you read more of local news or world news?
- When do you usually read news? Which part of the day suits best to consume news?
- How much time do you allocate to read news on a daily basis?
- Do you Prefer immersive news or just headlines? On which topics do you prefer immersive news and which topics do you prefer just the headlines?

Validated Hypothesis

- Users prefer to read online news rather than news papers.
- Users in India watch Sport News more than they watch movie news
- Users would like to have notification or pop ups for only highly emergency news.
- Users prefer flipping pages than scrolling down
- Users prefer to read news in batches whenever they are free during the day
- Users in mid 30s prefer immersive news for wealth management.

Hypothesis that was Proved Wrong

- Users actually prefer a bright and streamlined UI
- Users in India within the age of 20 to 45 actually prefer to read news in English rather than their regional languages
- World news is also given equal importance as compared to regional news

Part 5: Product Artefacts

NAME: Akshay Kumar DEMOGRAPHICS: Age 30 Years Bangalore IT Professional Married Income: 18 Lakhs		
BACKGROUND	MOTIVATIONS	GOALS
Akshay Kumar is an IT professional living in Bangalore with his wife. He has busy life and thus difficult to keep track of latest news. He uses news aggregator apps to curate information of his interests. Along with entertainment news in sports and bollywood he is now much interested in career and wealth management information.	<ol style="list-style-type: none"> 1. Fear of Missing Out(FOMO) on latest news 2. Socialize effectively with his network. 3. Secure fast and affordable way to read news 	<ol style="list-style-type: none"> 1. Stay uptodate with current affairs, 2. Information on wealth management 3. Use his knowledge as a tool to discuss, build relationships with clients and colleagues
PAIN POINTS FRUSTATIONS	TECHNOLOGY	BRANDS AND INFLUENCERS
<ol style="list-style-type: none"> 1. Repeated Ads while reading news 2. Not able to curate and go back to news articles read with ease 3. No personalized information content displayed 4. Some news needs more in depth information and some needs to be just aware of. This "classification" of short and long news not available 	User Windows OS Laptop with chrome and firefox browsers, Uses Android OS for his mobile.	<ol style="list-style-type: none"> 1. Cult Fit for fitness 2. Follows Youtube channels on sports and wealth management 3. Follows "Pranjwal Kamra" wealth management Influencer
	QUOTE: " I want to have some BAU news along with discoverable and trending news coming right into my mobile app with the flexibility of curating my news feed"	

NAME: Akshay Kumar DEMOGRAPHICS: Age 30 Years Bangalore IT Professional Married Income: 18 Lakhs		
USER GOALS	USER EXPECTATIONS	PROCESS
<ol style="list-style-type: none"> 1. Stay uptodate with current affairs, 2. Information on wealth management 3. Use his knowledge as a tool to discuss, build relationships with clients and colleagues 	<ol style="list-style-type: none"> 1. To have some BAU news along with discoverable and trending news coming fast right into mobile app with the flexibility of curating the news feed 2. To use information to socialize effectively with network. 	<ol style="list-style-type: none"> 1. User feels bore and wants to get entertained with news and media 2. Searches Google or gets bits of info from social media in random way 3. Retention rates of read news is low, Not able to curate information consumed in strategic way 4. User follows specific influencers via blogs, podcasts.
EXPERIENCE	TOUCH POINTS	PAIN POINTS
<ol style="list-style-type: none"> 1. Shortlisting Info via google and social media is possible but takes time, effort 2. No curated and personalized content served 3. Trust quotient is low as no trust worthy aggregator involved. 4. No implicit feedback captured which leads to loss of vital data 	<ol style="list-style-type: none"> 1. Market Place applications 2. Social Media Listings 3. Friends, Family, Colleagues 4. Influencer generated multimedia 	<ol style="list-style-type: none"> 1. Repeated Ads while reading news 2. No curated and personalized podcasts 3. No personalized information content displayed 4. Some news needs more in depth information and some needs to be just aware of. This "classification" of short and long news not available
	IDEAS: Have dedicated app to curate news & podcasts for the underlying user based on his local, international preferences, phase of life he is currently in, classify news into short and long reads and serve him based on his interests and attention span availability	

Part 6: Reimagine Your Product

Areas of Opportunity

1. Video Streaming experience should be improved by providing subtitles.
2. Podcast quality and its recommendations should improve
3. Better Implicit feedback mechanisms to be included to understand the user and their needs better.
4. Switch language for specific news articles, podcasts feature to be made available
5. Users should be allowed to save and share any item for future use.
6. Classify News as summarised text and immersive text using GPT Technologies

Initiatives that does not require AIML Solution:

1. Switch language for specific news articles, podcasts feature to be made available
2. Users should be allowed to save and share any item for future use.

Initiatives that require AIML Solution:

1. Podcast quality and its recommendations should improve
2. Video Streaming experience should be improved by providing subtitles.
3. Classify News as summarised text and immersive text using GPT Technologies and supervised learning classification paradigm. Later recommend the same to users based on their preferences.
4. Better Implicit feedback mechanisms to be included to understand the user and their needs better.

Justification of AIML Initiative:

1. Podcasts now make up a considerable share of all data floating around. AI can revolutionise it by clear recommendation engines.
2. We have both implicit and explicit feedback coming from users such as Most Played, Liked on Daily Hunt, News Episodes watched, Duration of Podcasts watched, Podcasts skipped, demographic information etc.
3. Increases User Retention Rates and Revenue

Feasibility of AIML Solution:

1. Daily Hunt already has collected user implicit and explicit user data. It needs to get used.
2. Amazon Personalize can be used for real time recommendations, personalized marketing campaigns etc and its scalable.
3. **Most Feasible Solution to Work On:** Recommendation engine to improve and personalize podcast quality

Part 7: Business Benefit Estimation

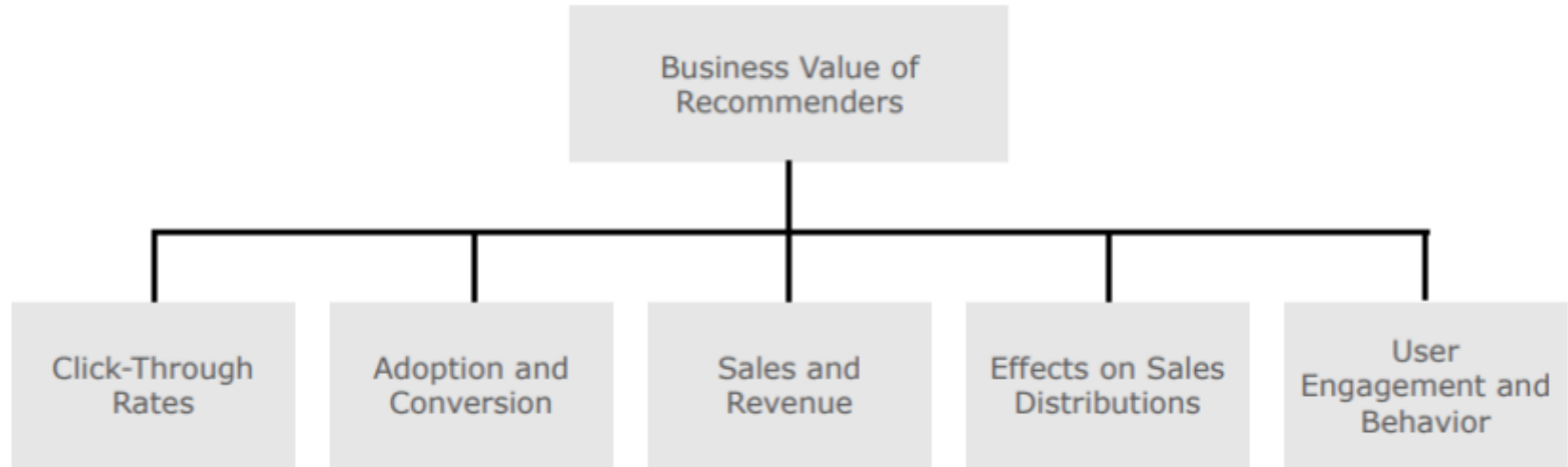
Predictive Accuracy Metrics: Predictive accuracy or rating prediction measures address the subject of how near a recommender's estimated ratings are to genuine user ratings. This sort of measure is widely used for evaluating non-binary ratings.

1. MAE: Mean Absolute Error
2. MSE: Means squared Error
3. RMSE: Root Mean Squared Error

We need to have low values of the above metrics for our predictive model to be a success.

Classification Accuracy Metrics: Classification accuracy measures attempt to evaluate a recommendation algorithm's successful decision-making capacity (SDMC). They are useful for user tasks such as identifying nice products since they assess the number of right and wrong classifications as relevant or irrelevant things generated by the recommender system.

Business Benefit Estimation- Business Success Metrics



Part 8: Data Setup

Type of Data: Recommendation Engine typically requires three types of data

- 1. User Data:** This data includes all data but not limited to demographic data such as age, gender, location, purchase history, search history and preferences. This is used to build a profile to make better recommendations.
- 2. Item Data:** This includes information about the items being recommended such as product descriptions, categories, tags, and ratings. Item data is used to understand the characteristics of the items and to find items that are similar to each other.
- 3. Contextual Data:** such as the time of day, weather, or other factors that may affect a users preferences and needs.

It is important to note that the quality of the recommendations generated by a recommendation engine is directly proportional to the quality and quantity of the data that goes into the model. Therefore, it is crucial to collect and maintain accurate and comprehensive data to ensure that the recommendation engine provides relevant and personalized recommendations.

Data for developing the AI/ML Solution-Sources of Data^{upGrad}

Sources of Data:

- 1.User-generated data:** This includes data that users create or provide, such as ratings, reviews, and feedback.
- 2.Transaction data:** This includes data about user purchases, searches, clicks, and other interactions with the system.
- 3.Social media data:** This includes data from social media platforms, such as user profiles, posts, and interactions.
- 4.Demographic data:** This includes information about users such as age, gender, location, and income.
- 5.Behavioral data:** This includes data about user behavior, such as time spent on the site, frequency of visits, and patterns of activity.

It's important to note that the quality of the data is critical to the effectiveness of the recommendation engine. The more accurate and diverse the data, the better the recommendations will be. Therefore, it's essential to gather data from a variety of sources and to continually update and refine the data to improve the recommendation engine's accuracy and usefulness.

Privacy Issues For Gathering the Data Needed:

1.User consent: Users should be informed about what data is being collected, how it will be used, and who will have access to it.

2.Data security: Data collected for building recommendation engine machine learning models needs to be stored securely to protect it from unauthorized access, theft, or misuse.

3.Anonymization: Personal information should be anonymized or pseudonymized before it is used for building recommendation engine machine learning models to prevent user identification.

4.Data sharing: If data is shared with third parties for building recommendation engine machine learning models, users should be informed about the parties involved, and their consent should be obtained.

5.Transparency: The process of data collection, storage, and usage should be transparent to users to build trust and to ensure that their privacy rights are respected.

Top 5 Features that could go into building the recommendation engine model:

1.Podcast metadata: This includes information about the podcast such as the title, description, genre, duration, and keywords. Podcast metadata can be used to understand the content of the podcast and to find similar podcasts.

2.User listening history: This includes data about the podcasts that a user has listened to in the past, how long they listened to each episode, and whether they subscribed or rated the podcast.

3.User demographic data: This includes information about the user such as age, gender, location, and interests. Demographic data can be used to segment users into different groups and to make recommendations based on those groups.

4.Social media data: This includes data from social media platforms, such as user profiles, posts, and interactions. Social media data can be used to understand the user's interests and to find podcasts that are relevant to those interests.

5.Collaborative filtering: This is a technique that involves analyzing the listening history and preferences of other users who have similar tastes to the current user. Collaborative filtering can be used to identify & recommend podcasts that are popular among similar users.

Part 9: Selection and Measure of Model

Identify Type of ML Problem and Justification

For Podcast personalizes feed we would need labelled training data as mentioned in the data setup in previous slides. Thus the **type of ML Problem is Supervised Learning** which can have both Regression and Classification. We use regression to predict a recommendation score or a relevance score of a particular item for a specific user. Also we can use Supervised classification method to classify as relevant or not relevant recommendation.

The corresponding metrics are “Predictive Accuracy metrics” and “Classification Accuracy Metrics”

Model Evaluation Metrics for Podcast recommendation engine:

- 1. Recall(Sensitivity):** Percentage of “Yes” correctly predicted by the model. Recall measures the proportion of podcasts that the user likes that are recommended by the recommendation engine. A high recall score indicates that the recommendation engine is able to capture the user's preferences effectively.
- 2. Precision:** Probability that a predicted “Yes” is actually a “Yes”. Precision measures the proportion of recommended podcasts that the user actually likes. A high precision score indicates that the recommendation engine is making accurate and relevant recommendations.
- 3. F1 score:** F1 score is the harmonic mean of precision and recall. It is a combined measure of both precision and recall and provides a balanced evaluation of the recommendation engine's performance.
- 4. Mean average precision (MAP):** MAP measures the average precision across all recommendations. It takes into account the order in which the recommendations are presented and penalizes the recommendation engine for presenting less relevant recommendations higher up in the list.
- 5. Root Mean Squared Error (RMSE):** RMSE measures the difference between the predicted and actual user ratings. It is used to evaluate the performance of recommendation engines that use explicit ratings, such as a user rating a podcast on a scale of 1 to 5.
- 6. Mean Absolute Error (MAE):** MAE measures the absolute difference between the predicted and actual user ratings. It is another evaluation metric that can be used for recommendation engines that use explicit ratings.

Each of these evaluation metrics provides insights into the accuracy, relevance, and effectiveness of the recommendation engine in providing personalized and relevant recommendations to users.

1.Recall(Sensitivity): Recall measures the proportion of podcasts that the user likes that are recommended by the recommendation engine. A high recall score indicates that the recommendation engine is able to capture the user's preferences effectively. This metric is important because it measures the system's ability to provide comprehensive recommendations that cover a user's entire range of preferences.

2.Precision: Precision measures the proportion of recommended podcasts that the user actually likes. A high precision score indicates that the recommendation engine is making accurate and relevant recommendations. This metric is important because it directly measures the accuracy of the recommendations made by the system.

3.F1 score: F1 score is the harmonic mean of precision and recall. It is a combined measure of both precision and recall and provides a balanced evaluation of the recommendation engine's performance. This metric is useful because it provides an overall evaluation of the recommendation engine's accuracy and comprehensiveness.

4.Mean average precision (MAP): MAP measures the average precision across all recommendations. It takes into account the order in which the recommendations are presented and penalizes the recommendation engine for presenting less relevant recommendations higher up in the list. This metric is important because it measures the quality of the ranking of recommendations.

5.Root Mean Squared Error (RMSE): & Mean Absolute Error (MAE): these metric is important because it measures the accuracy of the prediction made by the system.

Part 10: Minimum Viable Product

ASPECTS/FEATURES:

1. Quick Onboarding and Profile Setup
2. Allow Selection of multiple languages for news feeds
3. Allow users to save and share articles easily.
4. Podcast recommendations with better audio Quality
5. Enable news live streaming with closed captioning and transcripts.
6. Classify News as summarised text and immersive text using GPT Technologies

Must Have Features and Reason:

1. **Easy Onboarding:** If there is a lengthy onboarding process the churn rate increases. If user is tracked by a unique account the data of user preferences can help us better hooking the user and thus drive revenue
2. **Allow selection of multiple languages for news feeds:** If user is allowed to select multiple languages on a news app it allows them to track news of multiple regions simultaneously.
3. **Allow users to save and share articles:** This step can be a prerequisite implicit data collection feature where we know which type of articles the user has saved and shared. These features can go into our next ML model of recommendations. It also builds brand when we measure the number of shared articles as a metric.
4. **Podcast recommendations with better audio quality:** Podcast drives more than 30% of user traffic and having a podcast recommendation engine is definitely a must have feature as of now.

Nice to have Features and Reason:

- 1. Enable news live streaming with closed captioning and transcripts:** With this user gets to watch the live feed of any language and understand the content. However as of now, from a market perspective we yet to have high speed internet and market reach. Also with more data collected from news articles and podcasts we are yet in the process of building a user profile. This feature can be our next to have feature.
- 2. Classify News/Podcasts as summarised text and immersive text using GPT Technologies:** This a great to have feature, but as of now we are still experimenting and collecting data of users with various features that go into building models. Some of these features can than be used in knowing whether a user likes summarised info or immersive info. This feature can again be our next to have feature.



Disclaimer: All content and material on the UpGrad website is copyrighted material, either belonging to UpGrad or its bonafide contributors and is purely for the dissemination of education. You are permitted to access print and download extracts from this site purely for your own education only and on the following basis:

- You can download this document from the website for self-use only.
- Any copies of this document, in part or full, saved to disc or to any other storage medium may only be used for subsequent, self-viewing purposes or to print an individual extract or copy for non-commercial personal use only.
- Any further dissemination, distribution, reproduction, copying of the content of the document herein or the uploading thereof on other websites or use of content for any other commercial/unauthorized purposes in any way which could infringe the intellectual property rights of UpGrad or its contributors, is strictly prohibited.
- No graphics, images or photographs from any accompanying text in this document will be used separately for unauthorised purposes.
- No material in this document will be modified, adapted or altered in any way.
- No part of this document or UpGrad content may be reproduced or stored in any other web site or included in any public or private electronic retrieval system or service without UpGrad's prior written permission.
- Any rights not expressly granted in these terms are reserved.