

Disease Diagnosis Market Segmentation

Analysing the respective market in India using Segmentation analysis for **Disease Diagnosis**

by

Murari Muniprathap
Archana S
Padala Akhil Reddy
Ishmeet Kaur
Shashi Kumar Reddy

Problem Statement

During the time when Machine Learning and Deep Learning are booming so much, it is very important to understand that all this knowledge is not of any use if we can't apply it to different areas and impact the humanity.

This dataset will help you apply your existing knowledge to great use. Applying Knowledge to field of Medical Science and making the task of Physician easy is the main purpose of this dataset. This dataset has 132 parameters on which 42 different types of diseases can be predicted.

Due to big data progress in biomedical and healthcare communities, accurate study of medical data benefits early disease recognition, patient care and community services. When the quality of medical data is incomplete the exactness of study is reduced. Moreover, different regions exhibit unique appearances of certain regional diseases, which may result in weakening the prediction of disease outbreaks.

Abstract

The use of Machine Learning and Deep Learning in the field of medical science has the potential to impact humanity significantly. A dataset with 132 parameters that can predict 42 different types of diseases has been created to make the task of physicians easier. Accurate study of medical data is important for early disease recognition, patient care, and community services. However, incomplete medical data can reduce the exactness of study. Additionally, different regions may exhibit unique appearances of certain regional diseases, which may result in weakening the prediction of disease outbreaks. Therefore, there is a need to effectively segment the dataset to target specific healthcare providers based on their unique needs, budget, and patient population. Customization of the dataset may be required to meet the specific needs of each healthcare segment while ensuring scalability, interoperability, and adherence to regulatory requirements. The potential impact of this dataset is significant, but significant customization and investment may be required to effectively implement it in different healthcare providers.

Fermi Estimation

1. How many different market segments can this dataset be effectively applied to?
 - Let's assume there are around 3 major market segments: primary care physicians, specialists, and hospitals.
2. How many diseases can be accurately predicted using this dataset?
 - Let's assume there are 42 different types of diseases that can be predicted using this dataset.
3. How many medical professionals and institutions are there in each market segment?

- Let's assume there are around 50,000 primary care physicians, 10,000 specialists, and 5,000 hospitals.
4. How many patients can each medical professional or institution serve?
- Let's assume that primary care physicians can serve an average of 2,000 patients, specialists can serve an average of 500 patients, and hospitals can serve an average of 10,000 patients.
5. How much customization is required for each market segment to effectively use this dataset?
- Let's assume that around 30% customization is required for each market segment.

Data Pre-Processing

Import Libraries

```
import numpy as np
import pandas as pd
import seaborn as sb
from matplotlib import pyplot as plt
import seaborn as sns
import plotly.express as px
from sklearn.model_selection import train_test_split
```

Data Gathering

```
train = pd.read_csv(r"C:\Users\soora\Downloads\Training.csv")
test = pd.read_csv(r"C:\Users\soora\Downloads\Testing.csv")
print(train.head())
print(test.head())
```

train.head()													
	itching	skin_rash	nodal_skin_eruptions	continuous_sneezing	shivering	chills	joint_pain	stomach_pain	acidity	ulcers_on_tongue	...	scurrying	skin_peeli
0	1	1	1	0	0	0	0	0	0	0	...	0	0
1	0	1	1	0	0	0	0	0	0	0	...	0	0
2	1	0	1	0	0	0	0	0	0	0	...	0	0
3	1	1	0	0	0	0	0	0	0	0	...	0	0
4	1	1	1	0	0	0	0	0	0	0	...	0	0
5 rows × 134 columns													
test.head()													
	itching	skin_rash	nodal_skin_eruptions	continuous_sneezing	shivering	chills	joint_pain	stomach_pain	acidity	ulcers_on_tongue	...	blackheads	scurrin
0	1	1	1	0	0	0	0	0	0	0	...	0	0
1	0	0	0	1	1	1	0	0	0	0	...	0	0
2	0	0	0	0	0	0	0	1	1	1	...	0	0
3	1	0	0	0	0	0	0	0	0	0	...	0	0
4	1	1	0	0	0	0	0	1	0	0	...	0	0

Data Cleaning

```
train.isnull().sum()
```

```
[ ] train.isnull().sum()

itching          0
skin_rash        0
nodal_skin_eruptions  0
continuous_sneezing  0
shivering        0
...
blister          0
red_sore_around_nose  0
yellow_crust_ooze  0
prognosis        0
Unnamed: 133      4920
Length: 134, dtype: int64
```

```
train = train.drop('Unnamed: 133', axis=1)
```

```
[ ] train.isnull().sum()

itching          0
skin_rash        0
nodal_skin_eruptions  0
continuous_sneezing  0
shivering        0
..
inflammatory_nails  0
blister          0
red_sore_around_nose  0
yellow_crust_ooze  0
prognosis        0
Length: 133, dtype: int64
```

```
train.isnull().any()
```

```
itching          False
skin_rash        False
nodal_skin_eruptions  False
continuous_sneezing  False
shivering        False
...
inflammatory_nails  False
blister          False
red_sore_around_nose  False
yellow_crust_ooze  False
prognosis        False
Length: 133, dtype: bool
```

```
] train.duplicated()
```

```
] 0      False
   1      False
   2      False
   3      False
   4      False
   ...
4915    True
4916    True
4917    True
4918    True
4919    True
Length: 4920, dtype: bool
```

```
] # check if balanced
train['prognosis'].value_counts()
```

```
] Fungal infection          120
Hepatitis C                 120
Hepatitis E                 120
Alcoholic hepatitis         120
Tuberculosis                120
Common Cold                 120
Pneumonia                   120
Dimorphic hemorrhoids(piles) 120
Heart attack                120
Varicose veins              120
Hypothyroidism              120
Hyperthyroidism             120
Hypoglycemia                120
Osteoarthritis              120
Arthritis                   120
(vertigo) Parosymal         120
Acne                        120
Urinary tract infection     120
...                        ...
```

```
train.corr()
```

	itching	skin_rash	nodal_skin_eruptions	continuous_sneezing	shivering	chills	joint_pain	stomach_pain	acidity	ulcers_on_tongue	...	pus_fi
itching	1.000000	0.318158	0.326439	-0.086906	-0.059893	-0.175905	-0.160650	0.202850	-0.086906	-0.059893	...	
skin_rash	0.318158	1.000000	0.298143	-0.094786	-0.065324	-0.029324	0.171134	0.161784	-0.094786	-0.065324	...	
nodal_skin_eruptions	0.326439	0.298143	1.000000	-0.032566	-0.022444	-0.065917	-0.060200	-0.032566	-0.032566	-0.022444	...	
continuous_sneezing	-0.086906	-0.094786	-0.032566	1.000000	0.608981	0.446238	-0.087351	-0.047254	-0.047254	-0.032566	...	
shivering	-0.059893	-0.065324	-0.022444	0.608981	1.000000	0.295332	-0.060200	-0.032566	-0.032566	-0.022444	...	
...
small_dents_in_nails	-0.061573	0.331087	-0.023073	-0.033480	-0.023073	-0.067765	0.359845	-0.033480	-0.033480	-0.023073	...	
inflammatory_nails	-0.061573	0.331087	-0.023073	-0.033480	-0.023073	-0.067765	0.359845	-0.033480	-0.033480	-0.023073	...	
blister	-0.061573	0.331087	-0.023073	-0.033480	-0.023073	-0.067765	-0.061889	-0.033480	-0.033480	-0.023073	...	
red_sore_around_nose	-0.061573	0.331087	-0.023073	-0.033480	-0.023073	-0.067765	-0.061889	-0.033480	-0.033480	-0.023073	...	
yellow_crust_ooze	-0.061573	0.331087	-0.023073	-0.033480	-0.023073	-0.067765	-0.061889	-0.033480	-0.033480	-0.023073	...	

132 rows × 132 columns

Segment Extraction

Train Test Split

```
A = train[["prognosis"]] # diseases
B = train.drop(["prognosis"],axis=1) # symptoms
C = test.drop(["prognosis"],axis=1) # symptoms - testing
trainx, testx, trainy, testy = train_test_split(B,A,test_size=0.2)
```

Apply Machine Learning Algorithms

Random Forest

```
mod = RandomForestClassifier(n_estimators = 100,n_jobs = 5, criterion= 'entropy',random_state = 42)
mod = mod.fit(trainx,trainy.values.ravel())
pred = mod.predict(testx)
```

```
metrics.accuracy_score(testy, pred)
```

```
1.0
```

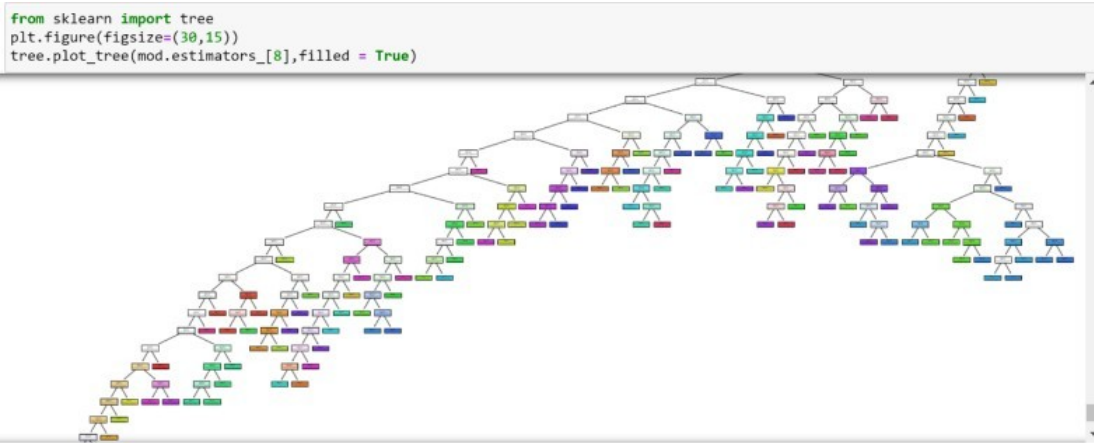
```
report = classification_report(testy, pred, output_dict=True)
pd.DataFrame(report).transpose()
```

	precision	recall	f1-score	support
(vertigo) Paroxysmal Positional Vertigo	1.0	1.0	1.0	28.0
AIDS	1.0	1.0	1.0	30.0
Acne	1.0	1.0	1.0	23.0
Alcoholic hepatitis	1.0	1.0	1.0	13.0
Allergy	1.0	1.0	1.0	15.0
Arthritis	1.0	1.0	1.0	25.0

```
test = test.join(pd.DataFrame(mod.predict(C),columns=["predicted"]))[["prognosis","predicted"]]
test['result']= ''
for i in range(len(test)):
    if test["prognosis"][i] == test["predicted"][i]:
        test['result'].iloc[i] = 'Correct'
    else:
        test['result'].iloc[i] = 'Incorrect'
test
```

	prognosis	predicted	result
0	Fungal infection	Fungal infection	Correct
1	Allergy	Allergy	Correct
2	GERD	GERD	Correct
3	Chronic cholestasis	Chronic cholestasis	Correct
4	Drug Reaction	Drug Reaction	Correct
5	Peptic ulcer disease	Peptic ulcer disease	Correct
6	AIDS	AIDS	Correct
7	Diabetes	Diabetes	Correct
8	Gastroenteritis	Gastroenteritis	Correct
9	Bronchial Asthma	Bronchial Asthma	Correct
10	Hypertension	Hypertension	Correct
11	Migraine	Migraine	Correct
12	Cervical spondylosis	Cervical spondylosis	Correct
13	Paralysis (brain hemorrhage)	Paralysis (brain hemorrhage)	Correct
14	Jaundice	Jaundice	Correct
15	Malaria	Malaria	Correct

```
cm = confusion_matrix(y_test, pred)
pd.DataFrame(cm)
```



Decision Tree

```
dt = DecisionTreeClassifier()
clf = dt.fit(trainx,trainy)
pred = clf.predict(testx)
scores = cross_val_score(clf, testx,testy, cv=5)
print("for Decision Tree: ")
print(scores.mean())
```

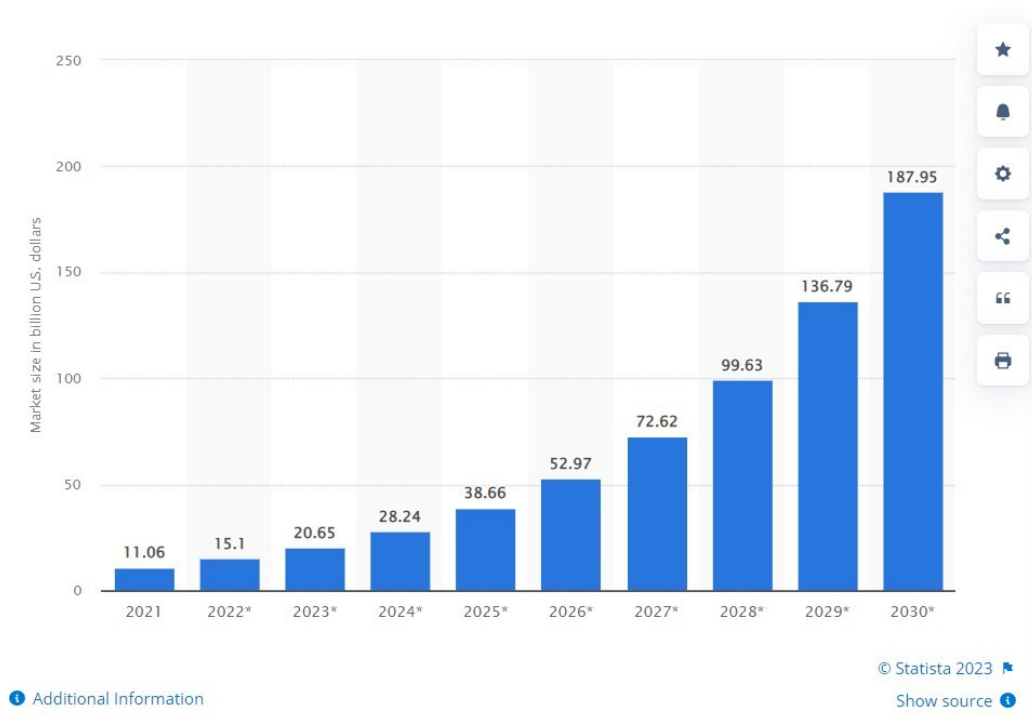
```
for Decision Tree:
0.9695224282606445
```

AI MARKET SHARE INDIA 2021, BY INDUSTRY PUBLISHED BY SHANGLIAO SUN, MAR 25, 2022

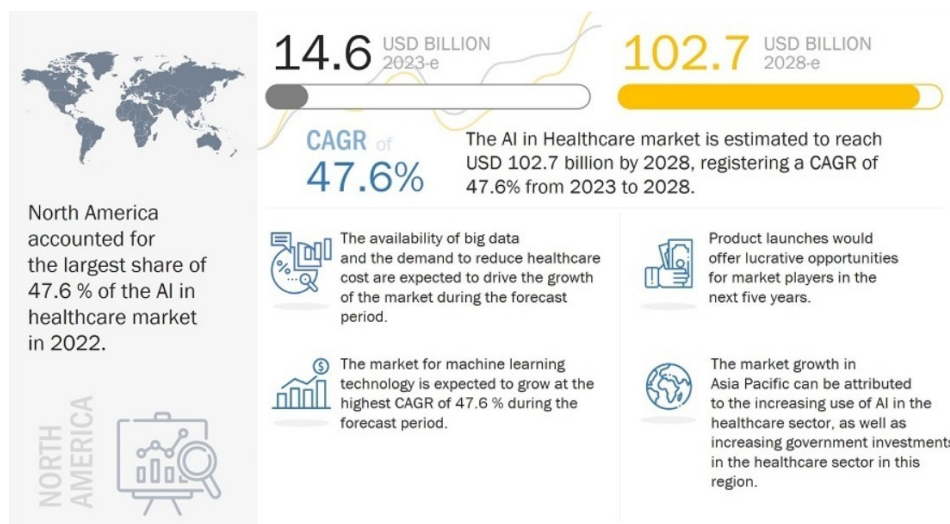
The AI market share of the IT services industry in India reached 51.8 percent in 2021. Artificial intelligence has been responsible for drastic changes in the technology sector where it can greatly improve productivity through process simplification and automation. It is also an integral part and one of the fundamental bases of Industry 4.0. In several developed countries, AI could potentially maximize labor productivity by more than 30 percent in the next 15 years.

- AI application in India** As India is a country with huge linguistic diversity, it imposes a great challenge to governments and companies when conducting business with people of different linguistic backgrounds. As a result, one of the first applications for AI in India is in the field of customer service. The Indian government has increased public investment to promote the Digital India initiative in the fields of AI, IoT, big data, machine learning, and robotics.
- Challenges of AI adoption in India** However, there are several obstacles India faces in the process of AI adoption. India has a comparatively small number of scientists and researchers in the field of machine learning and artificial intelligence. It also lacks sufficient qualified specialists to localize and implement the latest technologies in the field. However, the Ministry of Electronics and Information Technology, along with various industrial bodies have introduced several programs of personnel training and technical infrastructure building to lay the foundation for future AI development in India.

AI IN MEDICAL DIAGNOSIS MARKET SHARE:



This figure shows AI in medical diagnosis market share globally



OPPORTUNITIES IN MEDICAL DIAGNOSIS

- ♦ India has just 64 doctors available per 100,000 people compared to a global average of approximately 150 per 100,000. Primary health centres and sub-centres at the rural peripheries are often woefully understaffed and lack critical infrastructure to meet patient needs. This often translates to a lack of high-quality diagnostic services, **especially in rural India**, home to more than 70% of the population.
- ♦ Leveraging predictive analytics by using Artificial Intelligence (AI) for early detection can be a powerful tool for targeted public health interventions, especially in the context of limited healthcare capacity and delayed disease detection capabilities outside of urban centres. AI-enabled tools offer opportunities to bridge these inequities and reach AI maturity in the healthcare market in India, which is expected to reach \$372 billion this year.
- ♦ Multiple leading technology players and healthcare companies are developing AI solutions for applications in healthcare. Philips Healthcare (Netherlands), Agfa- Gevaert (Belgium), GE Company (US), and Siemens Healthineers (Germany) have started integrating AI into their medical imaging software systems. For instance, in collaboration with NVIDIA Corporation, GE has 500,000 imaging devices in use worldwide. These devices use AI to speed up the process of analyzing CT scans with improved accuracy. Siemens Healthineers' AI Rad Companion Chest CT is a software assistant that uses AI for CT.

ARTIFICIAL INTELLIGENCE IN HEALTHCARE MARKET HIGHLIGHTS:

The study categorizes the AI in medical diagnostics market into the following segments and sub-segments:

Segment	Subsegment
AI in Healthcare Market, by Offering:	<ul style="list-style-type: none">• Hardware<ul style="list-style-type: none">◦ Processor<ul style="list-style-type: none">▪ MPU▪ GPU▪ FPGA▪ ASIC◦ Memory◦ Network<ul style="list-style-type: none">▪ Adapter▪ Switch▪ Interconnect• Software<ul style="list-style-type: none">◦ AI Platform<ul style="list-style-type: none">▪ Application Program Interface (API)▪ Machine learning Framework◦ AI Solution<ul style="list-style-type: none">▪ On-Premise▪ Cloud• Services<ul style="list-style-type: none">◦ Deployment and Integration◦ Support & Maintenance

AI in Healthcare Market, by Technology:

- Machine Learning
 - Deep Learning
 - Supervised
 - Unsupervised
 - Reinforcement Learning
 - Others
- Natural Language Processing
 - IVR
 - OCR
 - Pattern and Image Recognition
 - Auto Coding
 - Classification and Categorization
 - Text Analytics
 - Speech Analytics
- Context-aware Computing
 - Device Context
 - User Context
 - Physical Context
- Computer Vision

AI in Healthcare Market, by End User:

- Hospitals & healthcare providers
- Healthcare Payers
- Pharmaceuticals & Biotechnology Companies
- Patients
- Others

AI in Healthcare Market, by Region:

- North America
 - US
 - Canada
 - Mexico
- Europe
 - UK
 - Germany
 - France
 - Italy
 - Spain
 - Rest of Europe
- APAC
 - China
 - Japan
 - India
 - South Korea
 - Rest of APAC
- RoW
 - Middle East & Africa
 - South America



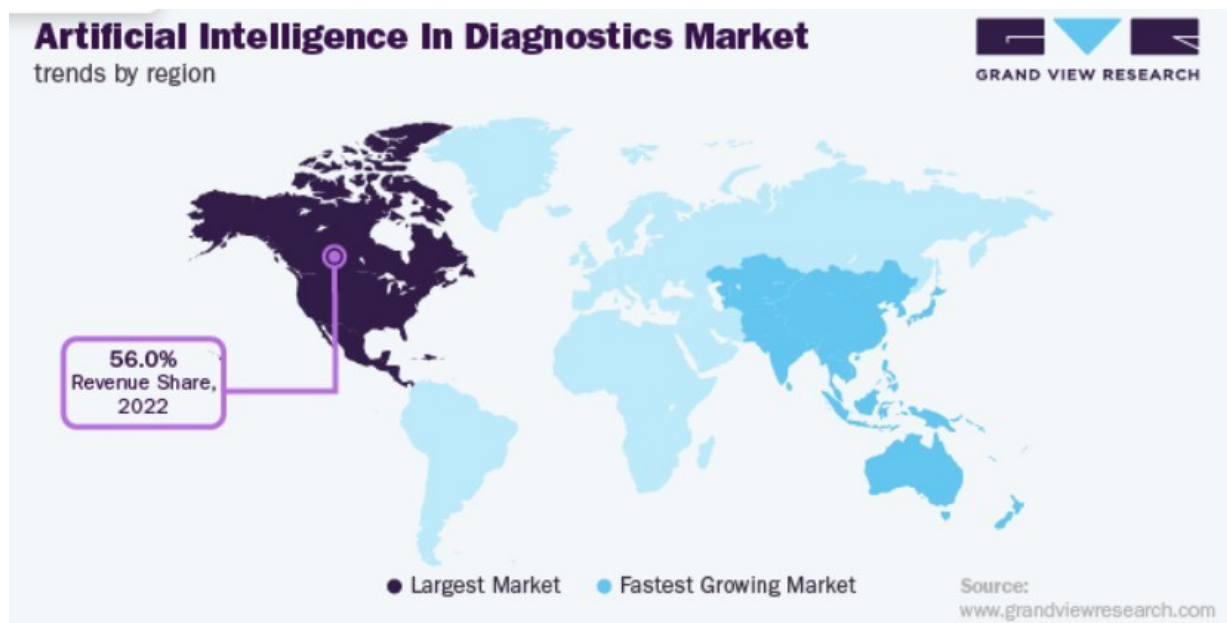
Source: Press Releases, Expert Interviews, and MarketsandMarkets Analysis

Artificial Intelligence in Healthcare Market by Region

- The application of AI in medical diagnostics is growing at a fast pace owing to factors such as rising government initiatives to drive the adoption of AI-based technologies, increasing adoption of AI solutions by radiologists to reduce work pressure, the influx of big data, availability of funding for AI-based startups, and the growing number of cross-industry partnerships & collaborations. However, the lack of a skilled AI workforce, ambiguity in regulations, and the reluctance among medical practitioners to adopt these solutions are factors expected to restrain the market growth.

Global AI in Medical Diagnostics Market Trends





AI IN MEDICAL DIAGNOSTICS MARKET DYNAMICS

- ♦ **Drivers: Influx of big data**

With the increasing digitization and adoption of information systems in the healthcare industry, big data (large and complex data) is generated at various stages of the care delivery process. In the medical diagnostics industry, big data comprises information generated from clickstream and web & social media interactions; readings from medical devices, such as sensors, ECGs, X-rays, healthcare claims, and other billing records; and biometric data, among other sources. Big data and analytical solutions have grown exponentially in sophistication and adoption in the last decade with the growing adoption of EHRs, digitized laboratory slides, and high-resolution radiology images among healthcare providers. Healthcare is one of the top five big data industries, especially in the US. In the coming years, the volume of big data in medical diagnostics is expected to increase due to the use of bidirectional patient portals, which allow patients to upload data and images to their EMRs. The need to efficiently manage the ever-increasing volume of large and complex medical diagnostic data is compelling the healthcare industry to turn its focus toward various AI-based solutions

- ♦ **Restraints: Reluctance among medical practitioners to adopt AI based technologies**

Extensive growth in digital health has enabled healthcare providers to assist patients through novel treatment approaches. AI technologies offer doctors tools that help them diagnose and effectively treat patients better. However, there is an observed reluctance among doctors in the acceptance of new technologies. For instance, medical practitioners have a misconception that AI will replace doctors in the coming years.

Doctors and radiologists believe that skills such as empathy and persuasion are human skills; thus, technologies cannot completely rule out the presence of a doctor.

Additionally, there is a concern that patients may be excessively inclined toward these technologies and may forgo necessary in-person treatments, which might also challenge long-term doctor-patient relationships. Several healthcare professionals have doubts about AI capabilities in accurately diagnosing patient conditions. Therefore, it is challenging to convince providers that AI-based solutions are cost-effective, efficient, and safe solutions that offer doctors convenience and better patient care. However, healthcare providers are increasingly accepting the potential benefits of AI-based solutions and the spectrum of applications it serves. Hence, there is a possibility that in the coming years, doctors and radiologists will be more inclined toward AI-based technologies in healthcare.

- ♦ **Challenges: Budgetary constraints**

Extensive growth in digital health has enabled healthcare providers to assist patients through novel treatment approaches. AI technologies offer doctors tools that help them diagnose and effectively treat patients better. However, there is an observed reluctance among doctors in the acceptance of new technologies. For instance, medical practitioners have a misconception that AI will replace doctors in the coming years.

Doctors and radiologists believe that skills such as empathy and persuasion are human skills; thus, technologies cannot completely rule out the presence of a doctor.

Additionally, there is a concern that patients may be excessively inclined toward these technologies and may forgo necessary in-person treatments, which might also challenge long-term doctor-patient relationships. Several healthcare professionals have doubts about AI capabilities in accurately diagnosing patient conditions. Therefore, it is challenging to convince providers that AI-based solutions are cost-effective, efficient, and safe solutions that offer doctors convenience and better patient care. However, healthcare providers are increasingly accepting the potential benefits of AI-based solutions and the spectrum of applications it serves. Hence, there is a possibility that in the coming years, doctors and radiologists will be more inclined toward AI-based technologies in healthcare.

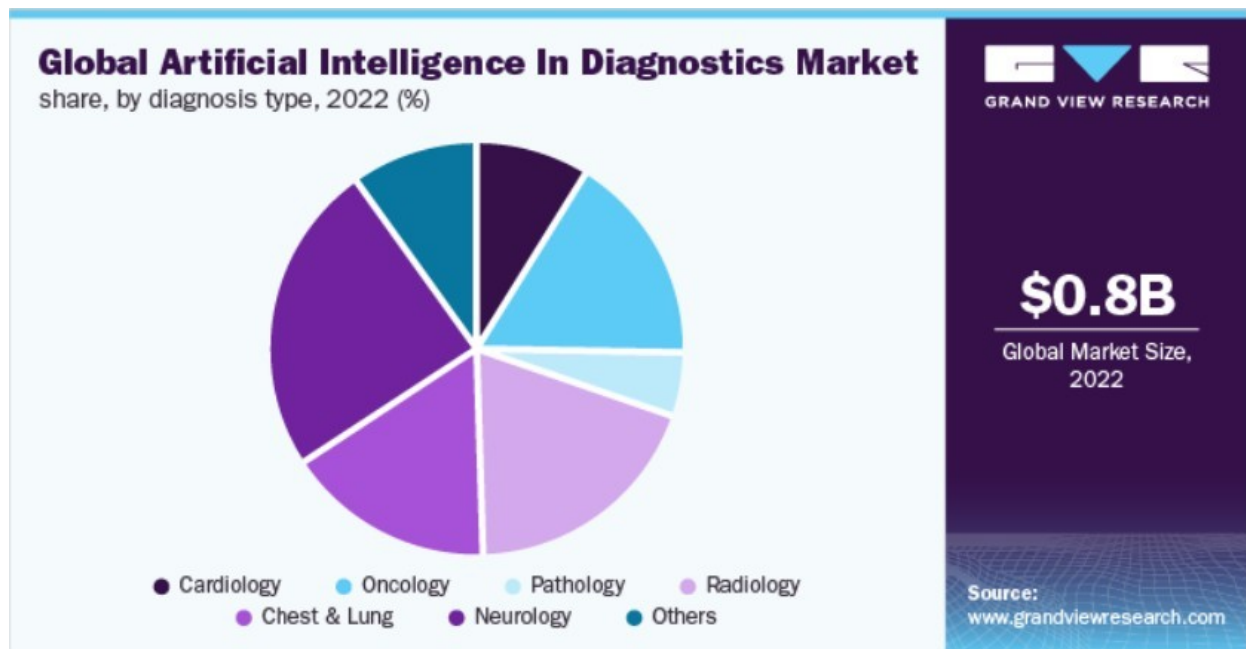
- ♦ **Opportunities: Increasing focus on developing human-aware AI systems**

The actual projections aimed during the emergence of AI technologies were to make them human-aware, i.e., developing models with the characteristics of human thinking. However, creating interactive and scalable machines remains a challenge for the developers of AI machines. Additionally, rise in human interference with AI techniques and interest to discover the machine learning process has introduced new research challenges, i.e., interpretation and presentation challenges such as issues with automating parts and intelligent control of



crowdsourcing. Interpretation challenges include challenges AI machines have in understanding human input, such as knowledge and specific directives. Presentation challenges include issues related to delivering the AI system's output and feedback mechanisms. Thus, developing human-aware AI systems remains the foremost opportunity for AI developers

Diagnosis Type Insights

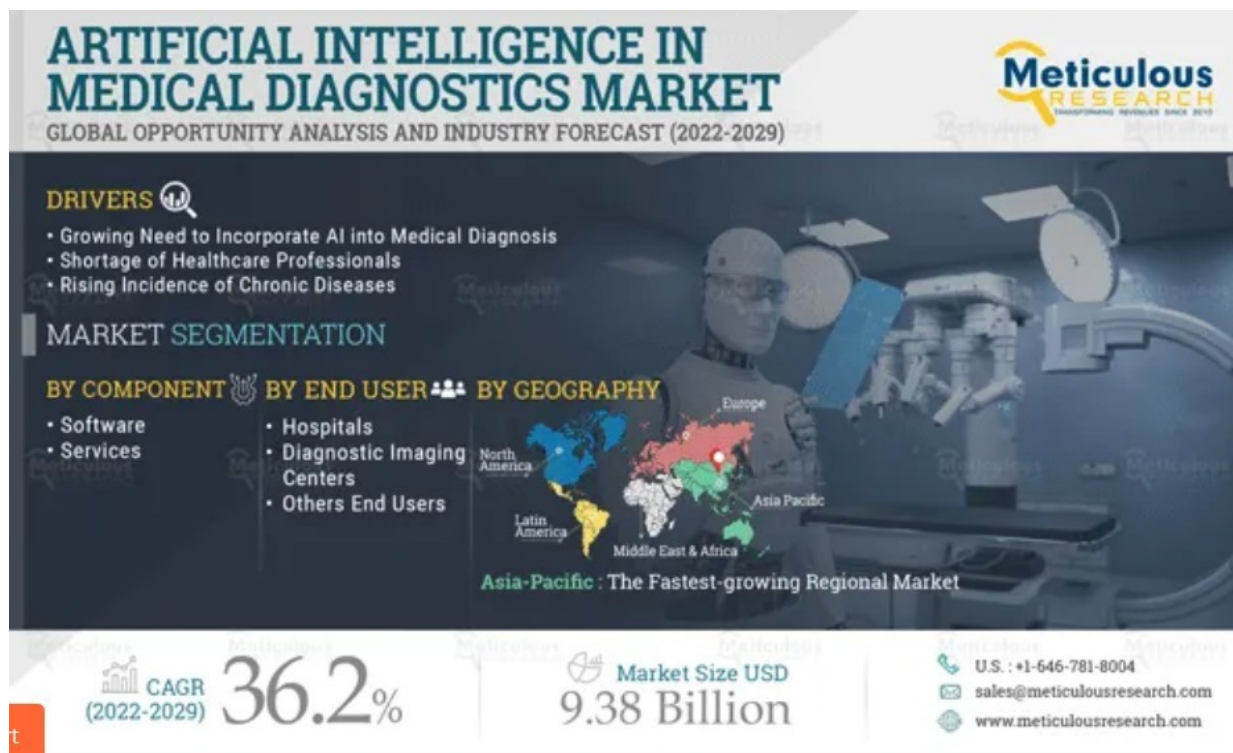


Global Artificial Intelligence In Diagnostics Market Segmentation

The Impact of COVID-19 on the AI in Medical Diagnostics Market

The outbreak of the COVID-19 pandemic adversely affected the healthcare sector globally. The number of people infected with COVID-19 skyrocketed, causing a huge burden on the health system worldwide. COVID-19 mainly affects the lungs of the patients. Hence, cardiothoracic imaging in COVID-19 cases is a common diagnostic practice to identify the severity of the disease. The number of research studies using AI techniques to diagnose COVID-19 rapidly increased in 2020. Many studies focused on diagnosing COVID-19 from chest CT images using AI technology. Several studies showed AI models to be as accurate as experienced radiologists in diagnosing COVID-19.





KEY FINDINGS OF THE AI IN MEDICAL DIAGNOSTICS MARKET STUDY:

The Software Segment to Dominate Market in 2022

In 2022, the software segment is expected to account for the larger share of the AI in medical diagnostics market. The large market share of this segment is attributed to the rising demand for AI-based software in diagnostics to deliver an accurate diagnosis in a short time, the rapid development of new AI algorithms and new software approvals, and applications of AI-based software in various fields, including radiology, cardiology, neurology, gynecology, and ophthalmology among others.

The Brain & Neurological Segment to Dominate Market in 2022

In 2022, the brain & neurological segment is expected to account for the largest share of the AI in medical diagnostics market. The large share of this segment is attributed to the benefits of AI-based solutions in diagnosing various neurological diseases, such as aiding radiologists in interpreting medical images to make a quick and accurate diagnosis, reducing noise in medical images, and creating high-quality images from lower doses of radiation.

The CT scan Segment to Dominate Market in 2022

In 2022, the CT scan segment is expected to account for the largest share of the AI in medical diagnostics market. The large market share of this segment is attributed to factors such as the growing need for effective and early diagnosis and the rising prevalence of chronic diseases globally due to the aging population. Furthermore, the outbreak of the COVID-19 pandemic increased the demand for chest CT examinations in COVID-19 patients, contributing to the market growth of this segment.

The Hospitals Segment to Dominate Market in 2022

In 2022, the hospitals segment is expected to account for the largest share of the AI in medical diagnostics market. The large share of this segment is attributed to factors such as the advantages of adopting AI-based solutions to automate the diagnosis process and reduce the workload in hospitals, the increasing number of patients undertaking diagnostic procedures, the growing need for early disease detection, and the shortage of medical specialists.

Asia-Pacific: the Fastest-growing Regional Market

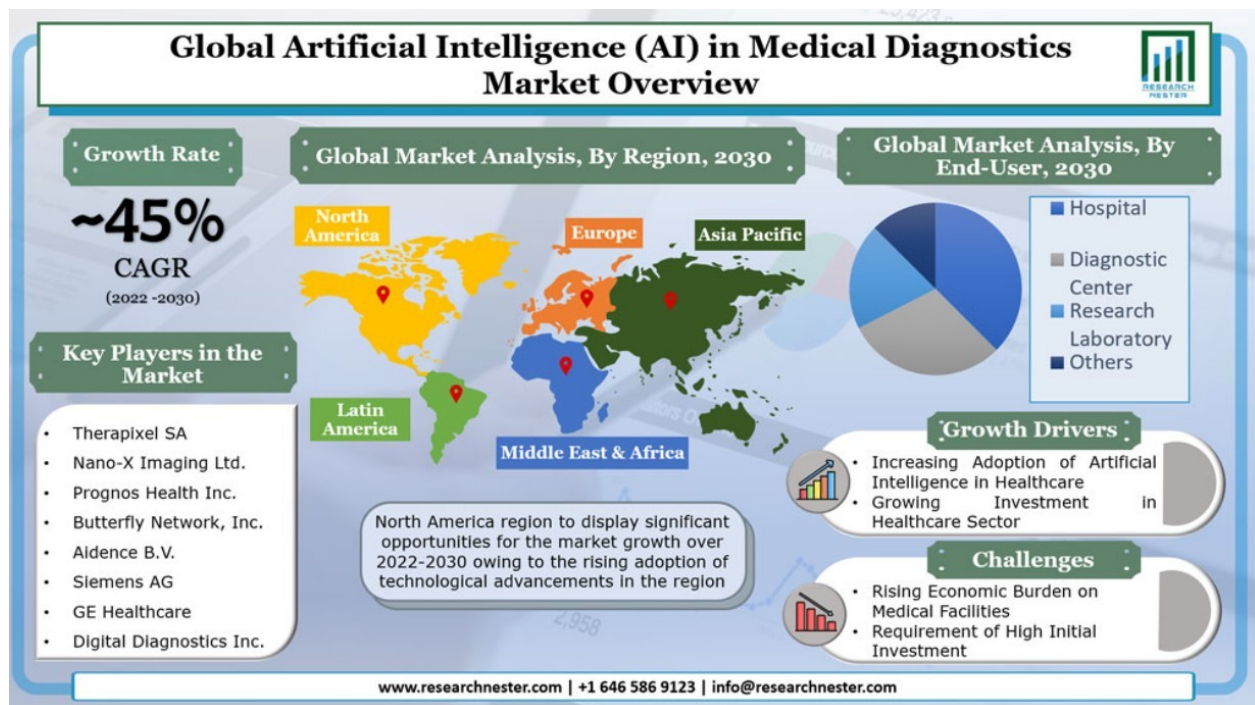
Asia-Pacific is slated to register the highest growth rate in the AI in medical diagnostics market during the forecast period. The high market growth in Asia-Pacific is attributed to the high growth opportunities due to the increasing prevalence of various chronic & infectious diseases, the increasing number of AI-based startups, especially in China and India, and the huge potential of AI in addressing the gap in the healthcare infrastructure in the region.

Key Players

The report includes a competitive landscape based on an extensive assessment of the market based on component, specialty, modality, end user, and geography, and the key growth strategies adopted by leading market players between 2019 and 2022.

The key players profiled in the AI in medical diagnostics market study are Siemens Healthineers AG (Germany), GE Healthcare (U.S.), Aidoc Medical Ltd. (Israel), International Business Machines Corporation (U.S.), AliveCor, Inc. (U.S.), VUNO Inc. (South Korea), Digital Diagnostics Inc. (U.S.), NovaSignal Corp. (U.S.), Riverain Technologies (U.S.), NANO-X IMAGING LTD (Israel), Imagen Technologies (U.S.), Koninklijke Philips N.V. (Netherlands), Agfa-Gevaert Group (Belgium), HeartFlow, Inc. (U.S.), and Arterys Inc. (U.S.).





The growth potential of AI in healthcare in India

AI expenditure in India increased by over 109% in 2018, totaling \$665 million and is expected to reach \$11.78 billion by 2025, adding \$1 trillion to India's economy by 2035.

NITI Aayog, a public policy think-tank linked to the Indian government, has been testing the application of AI in primary care for early detection of diabetes complications, and is currently validating the use of AI as a screening tool in eye care, by comparing its diagnostic accuracy with that of retina specialists. Integrating AI capabilities with portable screening devices, such as 3Nethra, can expand the capacity for eye screenings and early detection, and enable access in remote places across the country.

Similar applications are possible in oncology. Tata Medical Center and the Indian Institute of Technology recently launched India's first de-identified cancer image bank: the Comprehensive Archive of Imaging. AI-based tools can use high-quality de-identified images to enable machine learning models to detect biomarkers and improve outcomes for cancer research.

In cardiovascular healthcare, a major and somewhat unique challenge for India, Microsoft's AI Network for Healthcare and Apollo Hospitals are developing a machine learning model to better predict heart attack risk. Using clinical and lab data from over 400,000 patients, the AI solution can identify new risk factors and provide a heart risk score to patients without a detailed health check-up, enabling early disease detection.

India's way forward

AI maturity in health requires critical investments in the capacity of the workforce, data and infrastructure, governance and regulatory mechanisms, design and processes, partnerships and stakeholders as well as innovative business models.

Integrating AI into healthcare systems also requires an understanding of AI in national curricula for medical and public health students, both academic and practical.

Similarly, the Indian government will need to make appropriate investments in data infrastructure, such as interoperability, unified EMR and data stewardship. This is essential to build trust and long-term integration of AI into India's healthcare system.

The government must also invest in and build public-private partnerships across the healthcare industry to facilitate coordination between academia, government, industry, NGOs and patient advocacy organizations. They should scale governance and regulatory mechanisms to provide appropriate oversight for privacy, fairness and transparency.

NITI Aayog's National Strategy for AI prioritizes principles of privacy, ethics, security, fairness, transparency and accountability, as well as alignment with the rights afforded by the Indian Constitution. India is a founding member of the Global Partnership on AI alliance and has thus far adopted a measured approach for integration of AI, in keeping with ethical and responsible standards. These principles must be applied in practice as the technology scales.

The way in which AI systems are integrated, too, will be crucial. Human-in-the-loop and human-centric designs that empower healthcare staff to understand how a decision is made and how to incorporate this knowledge into treatment will minimize risk.

Investments in expanding the healthcare workforce and data literacy will build an informed workforce capable of leveraging AI in healthcare. India's measured adoption of this technology can enable it to bridge rural-urban disparities without leaving anyone behind, while becoming a leader among other emerging markets on the road to meeting the Sustainable Development Goals.

Summary

Extracted segments and profiling of potential segments:

Here is a brief description and profiling of each of the segments mentioned earlier in the AI in medical diagnosis market in India. Overall, these segments provide insights into the potential markets and opportunities for growth and investment in AI in medical diagnosis in India.:

1 Application:



- ♦ **Oncology:** The use of AI in oncology has great potential to improve cancer diagnosis, treatment planning, and patient outcomes. AI can analyze large amounts of medical imaging and patient data to identify patterns and predict treatment responses, enabling personalized treatment plans.
- ♦ **Neurology:** AI can help improve the accuracy and speed of diagnosis of neurological disorders, which are often difficult to diagnose and treat. AI can analyze brain images and patient data to identify patterns and provide insights into disease mechanisms and treatment responses.
- ♦ **Cardiology:** AI can help improve the accuracy of diagnosis and treatment of cardiovascular diseases, which are a leading cause of death worldwide. AI can analyze cardiac images and patient data to identify patterns and predict the risk of cardiovascular events, enabling personalized treatment plans.
- ♦ **Radiology:** AI can help improve the accuracy and efficiency of medical imaging diagnosis, which is a key component of medical diagnosis and treatment planning. AI can analyze medical images to identify patterns and abnormalities, enabling faster and more accurate diagnosis.
- ♦ **Others:** AI can be applied in other medical areas such as gastroenterology, dermatology, and ophthalmology to improve the accuracy and speed of diagnosis and treatment.

2 Technology:

- ♦ **Machine Learning:** Machine learning algorithms can analyze large amounts of medical data to identify patterns and predict treatment responses, enabling personalized treatment plans.
- ♦ **Deep Learning:** Deep learning neural networks can analyze complex medical data such as brain images to identify patterns and provide insights into disease mechanisms and treatment responses.
- ♦ **Natural Language Processing (NLP):** NLP can help analyze medical text and patient data to identify patterns and insights, enabling faster and more accurate diagnosis and treatment planning.
- ♦ **Others:** Other AI technologies such as computer vision and robotics can help improve the accuracy and efficiency of medical diagnosis and treatment.

3 End-User:

- ♦ **Hospitals:** Hospitals are the primary end-users of AI in medical diagnosis, as they use AI to improve the accuracy and speed of medical diagnosis and treatment planning.
- ♦ **Diagnostic Centers:** Diagnostic centers use AI to analyze medical images and provide diagnostic reports, enabling faster and more accurate diagnosis.



- ♦ **Research Institutes:** Research institutes use AI for medical research and drug discovery, enabling faster and more efficient drug development and personalized medicine.
- ♦ **Others:** Other end-users such as pharmaceutical companies and government agencies can use AI to improve the accuracy and efficiency of medical diagnosis and treatment.

4 Region:

- ♦ **North India:** The northern region of India has a large population and a high burden of chronic diseases such as cancer and cardiovascular disease, making it a potential market for AI in medical diagnosis.
- ♦ **South India:** The southern region of India has a well-developed healthcare infrastructure and a high level of medical expertise, making it a potential market for AI in medical diagnosis.
- ♦ **East India:** The eastern region of India has a large population and a high burden of infectious diseases such as tuberculosis, making it a potential market for AI in medical diagnosis.
- ♦ **West India:** The western region of India has a strong pharmaceutical industry and a high level of medical expertise, making it a potential market for AI in medical diagnosis and drug discovery.

The market is expected to grow significantly in the coming years due to the increasing demand for efficient and accurate diagnosis, rising prevalence of chronic diseases, and the need to reduce healthcare costs. Additionally, the government of India is also taking initiatives to promote the use of AI in healthcare, which is further expected to drive the market growth.

Customizing the Marketing Mix

The marketing mix is a critical tool for businesses to plan and execute their marketing strategies effectively. It consists of four major components - product, price, promotion, and place (distribution). These components must be designed to work together seamlessly to achieve a business's marketing objectives. In this task, we will explore how the given dataset can be used to customize the marketing mix for a pharmaceutical company.





Product

The first component of the marketing mix is the product. For a pharmaceutical company, the product is the drug or medicine they manufacture. The dataset provided can be used to customize the product in several ways. For instance, the company can focus on manufacturing drugs that target the most common symptoms and conditions listed in the dataset. For example, drugs that relieve symptoms of itching, skin rash, joint pain, stomach pain, acidity, vomiting, fatigue, cough, high fever, headache, abdominal pain, and diarrhea. The company can also focus on manufacturing drugs for conditions that are prevalent in the dataset, such as fungal infections, urinary tract infections, and liver failure.

Price

The second component of the marketing mix is the price. The pricing strategy for pharmaceutical companies is usually complicated due to various factors such as research and development costs, manufacturing costs, and regulatory requirements. However, the dataset can still be used to determine the pricing strategy for drugs based on their target market. For example, drugs that target common symptoms such as cough or fever may be priced affordably to appeal to a broader audience. In contrast, drugs for rarer conditions such as liver failure or fungal infections may be priced higher due to the higher costs associated with their production and distribution.

Promotion

The third component of the marketing mix is promotion. Promotion refers to the methods and channels used to promote the product to the target audience. In this case, the pharmaceutical company can use the dataset to design its promotional strategy. The company can target its promotions to areas with high incidences of the conditions listed in the dataset. For instance, they can advertise their liver failure drugs in areas with a high incidence of liver disease. The company



can also use social media platforms such as Facebook, Twitter, and Instagram to reach a broader audience.

Place (Distribution)

The fourth component of the marketing mix is place or distribution. This component focuses on how the product is delivered to the target market. For a pharmaceutical company, distribution is often complex due to regulatory requirements and the need for proper storage and transportation of drugs. However, the dataset can still be used to determine the distribution strategy. For example, drugs that target symptoms such as cough or fever can be distributed through retail pharmacies or drugstores, while drugs for rarer conditions such as liver failure or fungal infections may be distributed through hospitals or specialized clinics.

Conclusion

In conclusion, the given dataset can be used to customize the marketing mix for a pharmaceutical company. The company can use the data to tailor their product, pricing, promotion, and distribution strategies to target specific conditions and symptoms. By doing so, the company can effectively reach its target audience and achieve its marketing objectives. However, it is essential to note that the marketing mix is not a one-time task. It requires constant monitoring and adjustment to remain relevant and effective in the ever-changing market.

For Business Markets

In the world of business, one of the key factors that determine success is the ability to identify and capitalize on early market opportunities. For businesses operating in the business-to-business (B2B) space, this means identifying potential customer bases and calculating potential profits in the early market.



To begin with, it is important to define what we mean by the "early market". The early market refers to the initial stage of a new product or service launch, where the product or service is still relatively unknown and untested in the market. During this stage, the customer base is likely to be smaller and more niche, consisting of early adopters and innovators who are willing to take a risk on a new product or service.



Identifying the potential customer base in the early market is a crucial step in determining the potential for sales and profits. This involves researching the market and identifying the key players, such as industry leaders, competitors, and potential customers. It also involves analyzing market trends and identifying emerging needs and opportunities.

Once the potential customer base has been identified, the next step is to calculate the potential sale (profit) in the early market. This can be done by multiplying the potential customer base by your target price range. For example, if your target price range is \$100-\$150 per unit and your potential customer base is 500, your potential profit in the early market would be \$50,000-\$75,000.

Of course, it is important to keep in mind that these calculations are only estimates and there are many factors that can impact sales and profits in the early market. For example, the level of competition, the quality and features of the product or service, and the effectiveness of marketing and sales strategies can all play a role in determining success.

It is also important to consider the long-term potential of the product or service beyond the early market. While the early market may provide a good starting point for sales and profits, it is important to have a long-term strategy in place to continue to grow and expand the business.

In order to succeed in the early market, businesses must be agile and adaptable, able to respond quickly to changing market conditions and customer needs. This requires a strong focus on innovation, as well as a willingness to take risks and experiment with new ideas.

Overall, identifying potential customer bases and calculating potential profits in the early market is a key step in building a successful B2B business. By doing thorough research and analysis, and by remaining agile and adaptable, businesses can capitalize on early market opportunities and set themselves up for long-term success.

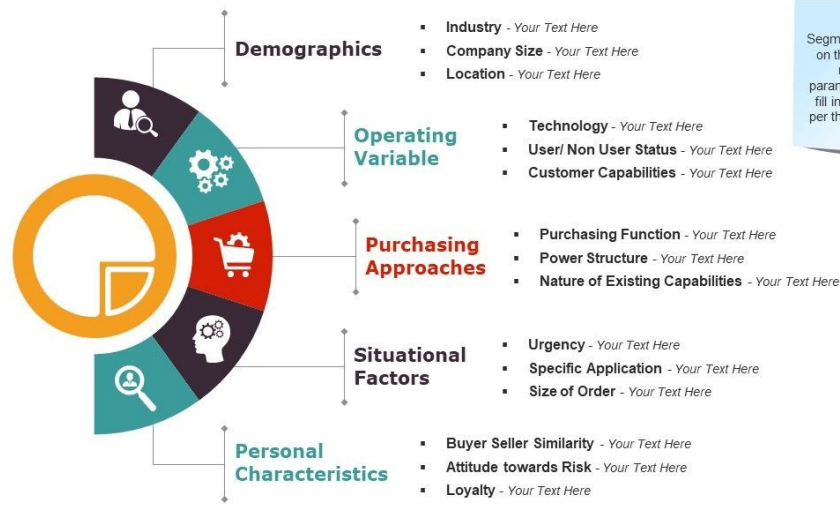
Market research and segmentation

Market research and segmentation are essential tools for businesses looking to understand their target audience and create effective marketing strategies. In the healthcare industry, understanding the symptoms and conditions that people experience can help healthcare providers better diagnose and treat patients. In this passage, we will discuss a wide range of symptoms and conditions, exploring their prevalence and impact on individuals.

Itching is a common symptom experienced by many people, and it can be caused by a variety of factors, including dry skin, insect bites, and allergies. Skin rash is another symptom that can be caused by numerous factors, including allergies, infections, and autoimmune disorders. Nodal skin eruptions are raised bumps on the skin that can be caused by infections, allergies, or other skin conditions.



Market Segmentation Business Markets



4

Continuous sneezing can be a symptom of allergies, colds, or other respiratory illnesses. Shivering and chills can be caused by a fever, infection, or exposure to cold temperatures. Joint pain is a common symptom of arthritis and other joint conditions, and stomach pain can be caused by a wide range of factors, including digestive issues, infections, and chronic conditions such as Crohn's disease and ulcerative colitis.

Acidity and ulcers on the tongue can be caused by acid reflux, while muscle wasting can be a symptom of diseases such as muscular dystrophy. Vomiting is a common symptom of many illnesses, including the flu, food poisoning, and stomach viruses. Burning micturition is a symptom that can indicate a urinary tract infection, and spotting during urination can be a sign of bladder or kidney issues.

Fatigue and weight gain can be caused by a wide range of factors, including hormonal imbalances, sleep disorders, and chronic conditions such as hypothyroidism. Anxiety and cold hands and feet can be symptoms of anxiety disorders, while mood swings and weight loss can be caused by depression or other mental health conditions. Restlessness and lethargy can be caused by a range of factors, including stress and fatigue.

Patches in the throat can be a symptom of strep throat or other infections, while irregular sugar levels can be an indicator of diabetes. Coughing is a common symptom of respiratory illnesses, including the common cold and flu. High fever is a symptom of many illnesses, including infections and autoimmune disorders, while sunken eyes and breathlessness can be symptoms of more severe illnesses such as pneumonia.

Sweating and dehydration can be caused by a range of factors, including exercise, hot temperatures, and illnesses such as heat stroke. Indigestion is a common symptom of digestive issues, including acid reflux and irritable bowel syndrome. Headache can be a symptom of many illnesses, including migraines and tension headaches, while yellowish skin and dark urine can be signs of liver issues.

Nausea and loss of appetite can be symptoms of a range of illnesses, including digestive issues and infections. Pain behind the eyes can be a symptom of sinusitis or other conditions, while back pain



can be caused by a range of factors, including poor posture, injury, and chronic conditions such as arthritis.

Constipation and abdominal pain can be caused by digestive issues, while diarrhea can be a symptom of infections and other conditions. Mild fever and yellow urine can be symptoms of a range of illnesses, while yellowing of the eyes can be a sign of liver issues. Acute liver failure is a severe condition that can be caused by a range of factors, including infections and alcohol abuse. Fluid overload can be caused by heart and kidney issues, while swelling of the stomach can be a sign of digestive issues such as bloating and constipation. Swollen lymph nodes can be caused by infections, autoimmune disorders, and cancer, while malaise is a general feeling of discomfort or unease that can be a symptom of many illnesses.

Business Modelling

AUDIENCE:

- Hospitals(both private and public), clinics run by specialists, and diagnostic laboratories which diagnose diseases of patients.

PROBLEM:

- Hospitals receive several patients everyday who come to get diagnosed and treated for various diseases.
- Often there would be long queues of patients and the waiting time would be very long. Sometimes this can pose a risk to some patients because they have to be diagnosed and treated in a short period of time. Some cases require emergency treatment compared to other cases.
- Contagious diseases can spread among patients who are waiting for a long time.

OFFERINGS:

The product offers automatic diagnosis at the hospitals which takes less time compared to in-person diagnosis by a doctor. This will reduce the waiting time significantly and more attention can be given to patients with severe medical conditions and emergency cases.

NEEDS:

- Data is needed on the symptoms of each disease, and for several diseases which cannot be identified merely by their symptoms, we need to collect data on the results of a large number of urine tests, stool tests and blood tests.
- We will also need images from X-ray, ultrasound and other medical imaging diagnostic techniques.
- Specialists are required to correctly label the collected data.
- The dataset has to be regularly updated.



- Data must be of good quality so that the model predictions will be accurate.
- Maintenance of the web application has to be done.

KEY PARTNERS:

- Hospitals.
- Testing Laboratories.
- Specialist Clinics, like dental clinics, dermatology clinics.

MONETISATION SCHEME:

- The best suited model for the product would be the subscription-based model.
- In this model, access to the product will be provided for a fixed amount of time for a fixed amount of fee.

FINANCIAL MODELLING:

For financial modelling, we need to consider the following variables:

- Number of small and mid-size hospitals which have subscribed to the product: $x_1(t)$
- Number of large hospitals which have subscribed to the product: $x_2(t)$
- Subscription cost for small to mid-size hospitals: Rs. 50000 per year per hospital.
- Subscription cost for large hospitals: Rs. 100000 per year per hospital.
- Cost of data collection = Rs. D lakh/year.
- Our team will have 2 ML engineers, 2 full-stack developers and 1 product manager.
- Cost to company for each ML engineer: 15 lakh per annum.
- Cost to company for each developer: 15 lakh per annum.
- Cost to company for the product manager: 20 lakh per annum.
- Total cost to company for the team: 90 lakh per annum.
- Therefore, the financial equation for the product will be:

$$y = 50000 * x_1(t) + 100000 * x_2(t) - D * 100000 - 9000000$$

- This equation doesn't take into account the initial one-time costs required to set the product, for example we have to collect a large amount of data in the beginning and there is a large cost associated with it, but it is done only once. This financial equation is for calculating the yearly profits.

GitHub Link

https://github.com/muni-prathap-goud/Disease_Diagnoses_Market_segmentation



