1. COMPACE and contrast data Mining w/ traditional statistical analy Nhat are the key differences of similarities  Diffinition  Traditional statistical analycu  I in hypothesis -driven and deals with statistical data that is stitle bears of statistical analycu  I in hypothesis -driven and deals with statistical data that is stitle bears of statistical data data data data data data data da	-5.5	B.517 2	ELECTIVE 2	Name: XIAN HUI B. CHEN
Nhat are the key difference; & diminarities  DEFINITION!  Traditional statistical analyses  In hypothesis -driven and deal with statistical data that is stated to be able to trovide symmaty & insight, by examining historical data (sales, reports, submer ratings, market trends  I myloty sechniques like regression analysis, hypothesis tes and probability distributions  Data Mining  I overall fraces of identifying rations of getting useful miles from big data sets. Instead of stortion hypothesis, it employs machine learning algorithms  I ministration methods aim to uncover relationshirs within data to into deution making. They both share the same goal, using talaniaves analyze & interprer information to define trends  I preferences  I report of information of define the data sets in ministrations of interpret information of define the data sets of indications of interpret information of define the data sets of indications of interpret informations of interpret informat				
DEFINITION  TRAditional statistical analycus  I hypothesis -driven and deals with statistical data that is still top to be a like topical symmathy of insights by examining historical data (sales, reports, costomer ratings, market trends)  I purply techniques like regression analysis, Mystothesis tes and pabability distributions  Data Mining  I rom big data sets  I rom big data sets  I often used on large unstructured data sets instead of startom hystothesis, it employs machine learning algorithms  I similarises  I entropies information to define the same goal, using tessinatives anonge & interpret information to define then ds  I preferences  I preferences  I productive of the same	1515	T STATISTICAL ANALYSIS		
Traditional statistical analysis  In hypothesis -driven and deals mith chatistical data that is set to be all to provide symmath & instatistical data that is set to be all to provide symmath & instatistical data that is set to be all to provide symmath & instance of the provide symmath of the provide symmath of the set of the se			- q Jimilanijiez	
Ly li hypothesis -driven and deals with staffstical data that is still be add in the provide summary & insight, by examining historical data (sales, reports, costomer ratings, market trends)  Ly Employs techniques like regression analysis, Myrothesis tes and probability distributions  Data Mining  Ly Overall process of Identifying patterns & getting useful insight from big data sets  Ly often used on large unstructured data sets instead of startion hyrothesis, it employs machine learning algorithms  Ly Both methods aim to uncover relationshirs within data to insome deution making. They both share the same goal, using taking vesting analyze & interpret information to define trends  Ly interpret information define trends  Ly interpret information define trends  Ly interpret information to define trends  Ly interpret information define trends  Ly interpret information defined tren				
beal 11 to Frovide Symmail & Insights by examining historical data (sales, reports, coshmer ratings, market trends  be employed techniques like regression analysis, hyrotherist test and probability distributions  pata mining  boverall process of identifying patterns & getting useful miss from big data sets  boften used on large unstructured data sets instead of starting hyrothesis, it employs machine learning algorithms  betting the methods aim to uncover relationships within data to inform deution making. They both share the same goal, using taking vesting analyze & interpret information to define trends  preferences  Transitional statistical pata mining  Approach hyrothesis driven (deductive) exploratory (inductive deases statistical stage, visitation data as a small dataset large, visitation data as a small dataset large & big data a stage stage and a small - medium dataset large & big data				
data (sale), reports, whemer ratings, market trends  Employs techniques like regression analysis, Myrothesis tes  and pobability distributions  Data Mining  Doverall process of Identifying patterns & getting use ful initial of starting high data sets  Trom big data sets  Trom big data sets  Trom big data sets  Trom big data sets  Diften used on large unstructured data sets instead of starting hyrothesis, it employs machine learning algorithm  Almitherites  The both methods aim to uncover relationshirs within data to inform the deution making. They both share the same goal, using takiniques analyze & interpret information to define trends  TRADITORNAL STATUTICAL ANALYSIS  APProach hyrothesis driven (deductive) exploratory (inductive deal set)  Dotaset structured small dataset large, unstructured seconds				
Employs techniques like regression analysis, Myrothesis tes  and employs distributions  pata Mining  Doverall emets of identifying ratterns & getting useful insignation of the most of th	al	examining historical	wait & invigits by e	- Goal U to Provide
Data Mining  Dota Mining  Doverall process of identifying ratterns & getting useful insignation big data sets  Soften used on large unstructured data sets instead of starting hyrothesis, it employs machine learning algorithms    Similarites    Doth methods aim to uncover relationshirs within data to into deusion making. They both share the same goal, using techniques analyze & interpret information to define trends    IRADITONAL STATISTICAL ANALYSIS    Approach hyrothesis driven (deductive) exploratory (inductive desired of the same set in		trends	omer ratings, market	data (sales, cepo
Data Mining  Doverall process of Identifying patterns & getting useful insist from big data sets  Trom big data sets  Doften used on large unstructured data sets instead of starting hyrothesis, it employs machine learning algorithms    Climicarities   Death methods aim to uncover relationships within data to information making. They both share the same goal, using techniques analyze & interpret information to define trends    Doffekenits   Approach   DATA MINING    Approach   DATA MINING    Dofaset   Structured Small dataset   large yunstructured    Scale   Small - medium dataset   large & big data a	11109,	ISIS, Myrothesis testin	ite regression analy	- Employs techn
Doverall process of Identifying patterns & getting useful insignation big data sets.  Trom big data sets			1001	and probability
Trom big data sets  Soften used on large unstructured data sets. instead of starting hypothesis, it employs machine learning algorithm    almitarites    Both methods aim to uncover relationships within data to information making. They both share the same goal, using techniques analyze & interpret information to define trends    DIFFERENCES    TRADITIONAL STATISTICAL ANALYSIS DATA MINING APProach hypothesis driven (deductive) exploratory (inductive dataset structured small dataset large, unstructured scale small - medium dataset large & big data a		4/39	·	pata Mining
Soften used on large unstructured data sets. instead of storting hyrothesis, it employs machine learning algorithms    State   State   State   State	9471	getting use ful mights	ntlfying ratterns &	- Overall Express
hyrothesis, it employs machine learning algorithm    almitarites    Both methods aim to uncover (elationshirs within data to info   deusion making. They both share the same goal, using techniques   analyze & interpret information to define trends   analyze & interpret information to define trends   TRADITIONAL STATISTICAL   DATA MINING   Approach   hyrothesis driven (deductive) exploratory (inductive dataset   large, unstructured     Structured, small dataset   large & big data a		NO AND LONG WAS A	SET I'M SAFETY BALLS	from big data set
DIFFERENCES  APProach  Approach  Detain Making  Traditional Statistical  Approach  Detain Making  Approach  Mypothesis driven (deductive)  Structured, small dataset  Structured and a medium dataset  Large & big data a	9 W/	. instead of starting w	nstructured data sets.	> often wed on 1
Deusion making. They both share the same goal, using techniques analyze & interpret information to define trends    DIFFERENCES    TRADITIONAL STATISTICAL ANALYSIS PATA MINING ANALYSIS driven (deductive) exploratory (inductive desired and seasons small dataset large, unstructured scale small - medium dataset large & big data a	1969/1/2	water Williams and Emily	ine learning algorithm	hyrothesis, It emplo
Detain methods aim to uncover relationships within data to info  Jeusian making. They both share the same goal, using techniques  analyze & interpret information to define trends    DIFFERENCES   TRADITIONAL STATISTICAL ANALYSIS  Approach  Nypothesis driven (deductive)  Exploratory (inductive)  Data set  Structured small dataset  Jarge & big data a	le kil. N	i. 7 Met View galaka	The Exemple on 1 199 mg	LIMILARITIES
APProach  Data Structured . Small dataset  Structured . Small dataset  Jerge & big data a	in	within data to inform	cover relationshirs	·
APProach  Data set  Structured . Small dataset  Scale  Analysis  Analysis  Approach  Structured . Small dataset  Stage & big data a				
APProach  ANAMSIV  ANAMSIV  PATA MININE  ANAMSIV  ANAMSIV  PATA MININE  Corporatory (inductive)  Dotasea  Structured, small datasea  Scale  Small - medium datasea  large & big data a				
APProach  ANAMSIV  ANAMSIV  ANAMSIV  ANAMSIV  ANAMSIV  ANAMSIV  ANAMSIV  ANAMSIV  ANAMSIV  Anama  An			wind the state of	LOVEE ERENIA (
Approach hypothesis driven (deductive) exploratory (inductive Dotaset Structured, small dataset large, unstructured scale small - medium dataset large & big data a	5	DATA MINING		T VIII CACINOS
structured, small dataset large, unstructured scale small - medium dataset large à big daja a	ve)	exploratory (inductive)		Approach
scale small-medium dataset large & big dasa a		large , unstructured date		
Automates mostly Manual highly automated	ipplications			
חסוןשיחופרעם ווייאווייון אווייון חסוןשיחפרעם	al-mortal	mostly manual highly automated		Automation
hypothesis reavised before analysis not required, find path	terns automatic	not required, find patterns	ed before analysis	hypothesis
	anous data types	night adaptable to vanor	(enongmules tom	
compitatinged poner lower, requires statistical higher, requires comp	erning	higher, required computy	requires statustical	compitational poner
FOIMULAY POWER		Power	formular	
THE RESERVE OF THE PROPERTY OF				

VALIANT

DATE	

2. Deveribe a specific data mining offlications in a field of	your
anothe Explain the problem being addressed, me data used, tech	
and the benefits achieved	
PCOblem	
Tuberculosis is one of the leading cause of morbidity in the	Chilles es
According to DAH & world Health Organization (with), the Ph	Heernes
Is among the for countries with the highest buiden or TB cas	es.
Early detection & timely intervention are used to control see	ad of
disease but our problem is low under reporting and delayed dia	
persiut .	
Tuolupon	-1-7-4-4
bata mining has been applied in the PH healthcare systed, part	beularly
in predictive TB diagnosis, to improve early detection rates	
oramize healthcare resources. Given the large volume of panent	
In Public hospitals of centers, data mining was unlined	
Data used	111
carge volume of patien records was used & patient data w/ the f	4 Where
collected:	
-> demographics (age, gender, location, socio economic background)	Maria :
-) medical history ( previous clinical illness, TB hubry )	ram T. Eta I. William
-> cilinical symptoms ( fever, cough, etc)	A Planta
- diagnostic test results	
-> I HEL HILL FORTON	
Techniques Applied	Latter to at
-> pension trees where a classification algorithm used to analyz	e paten;
records & Symptoms to determine whether a person u high rule-in Te	
- Arthoual Neural Networks was used in X-Ray images to defect TB relate	
abnormalities, reducing the need for radiologus in remote aree as	
-> Matural language Processing (NLB) -> extracts patient information	1 Enn
unstrudired dinical notes in electronic medical rooms	(141)
Example DOH's TB innovations & health system strengthening proje	
Andrew Land A seem of seem described by	VALIANT

DATE			

Benefits	DATE
	ms have made diagnosis accessible to
more filipinas. It also helps	in efficient resource allocation since by
	11 Dolf can allocat testing resources to
	reduce diagnishe errors & automate th
epacess	and the second of the second o
3. Explain the importance of data	a preparation in data mining, particularly
concerning privacy & security p	process. What die some common data
preprocessing techniques & w	thy are they necessary
-> IMPORTANCE	commended to all expense of lawrence in
	ata & raw data is often incomplete,
	lithout preparation, data mining modely
	The state of the s
can be maccurate.	a dada taran a traja da aran manda a sa la a
Trechmont	
Data Cleaning Data Integra	non pata mansso mation Data Reduction
V 12 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	wall well and the state of the
remove duplicate merging date	and to take and to the land
or incorrect entries Sources	Standardation then in Formation my
	be simplified
4. biscuss the ethical consideration	s surrounding data mining, process what
	privacy & seconty. Whey measures can
	cens of many and and an an an analysis and an an analysis and
	cerns in data mining revolve aroud privacy,
	suse of Personal data. Since many soco
	the users, some of them without explicit
consent, this raises concerns	
	erienced a data breach, hacking & exposed
	mation. All the sensitive data lite name, address
	were xeated and was publicly agessed online which
puts Filipino's of rux of tiden	they there & fraud

[MITI GATION] - All the measures in mitigating this revolves around

**CS** CamScanner

VALIANT

DATE
mplementing strong data protetion:
1) compliance w/ Data Privacy Laws - Data Privacy Act 2012 ensules
responsible data collection.
1 Data Encyption & Anony mization - to protect data, we need to encrypt
them especially sensitive data, by doing this it ensures protetion when
analyzing large dataset.
) user consent & Transparency - Before collecting the data, companies whould
as k for concent & communicate how it will be used.
Accountability & Ameal Al Development -> Al & ML models whould be
designed to avoid biases & discrimination in decuron-making
. How can biases in data affect thei results in data mining models? what
steps con be taken to identify & mitigale bias in data mining?
- Bias in data mining distorts outcomes which reads to incorrect conclusions.
Bias comes from unbalanced datasets or selection bias.
= Example in hiring algorithms, if training data consists of past hiring
decisions that favored a particular demographic, the hiring model
may continue to exhibit discriminatory behavior.
and the second of the second o
MITIGATION
by organizations must ensure data diversity and regularly audit models
+ techniques like re-sampling and re-weighting balances the data,
what it does is creating new samples based on one observe "ample
s use explainable AI (XAI) techniques to improve transparency in
deusion-making. This is a field of Al re-search that helps humans
understan how Al systems make decision
Choose one data mining algorithm. Briefly explain how it works,
what types of problem its suited for, and its strengths & weatness

Spensio Tree - wa popular data mining algorithm used for

class i fication & regression

La it solves problems through hierarchical tree structure with one
top call not node & branches out to different possible outcomes.
ADDI Node   -> entile dataset
based on internal Internal
the input ->   Node   Node
peqtures
Leaf   Leaf
Node Node Node Node
G terminal nodes at the end of branch - outcomes / Predictions
example: Deciding it you're threat the tree suggest if you need to don't coffee
Time of Day
morning afternoon
morning afternoon
Tried No
Tred? [No]
[Yes] No
Ves No
Conce The
Dimk Np
weree
Teva anal 1
Texpianation
Tree checks time of day (morning / aftern oon)
↓
ast if You're tired
11 tired, it suggests dink coffee
it it's afternoon it asks again if You're tred
I final condusion is made based on the arm er



