MID REVIEW PRESENTATION
ON

‘Developing materials informatics systems for alloy design’

Presented By:
1.Mansi Srivastava(5000XXXX)
2.Bhavy Mutreja(5000XXXX)

Under the Guidance of:
Mr. Dishant Beniwal

UPES
UNIVERSITY WITH A PURPOSE

UPES
UNIVERSITY WITH A PURPOSE

ANNEXURE X

Department of Mechanical Engineering
Project objective finalization (SYNOPSIS)

Date of Evaluation: 30/10/2020			
Program	B.Tech Mechanical	Semester	7
Mentor	Mr. Dishant Beniwal	Project ID	
Project title	Developing materials informatics systems for alloy design		

S.N	Criteria	Remark	Feasible/not Feasible
1	Relevant to the program and course outcome		
2	Methodology		
3	Cost estimate and funding source		
4	Expected time chart		
5	Availability of resources		

	Objectives Finalized	Approved/ Not approved
Objective 1	Create an extensive alloy database from existing literature.	
Objective 2	Implementing machine learning algorithms for structure and property prediction.	
Objective 3	Model validation through comparison with reported/experimental results.	

*Group members from different programs/ school or Faculty mentors from different programs/ school

Multi-disciplinary*	
Thrust Area relevant	
Research publication/ Conference/ Article	
Product Development/ Patenting	

Students Details		
Name	SAP-ID	Department
Mansi Srivastva	XXXX	Mechanical
Bhavy Mutreja	XXXX	Mechanical

UPES

Note 1: Attach synopsis with cost estimation and time chart. Without this the project will be rejected

Informatics system: Predicting the critical Temperature of Superconductors

- Informatics system : Informatics is the study of the structure, behavior, and interactions of natural and engineered computational systems. Understanding informational phenomena - such as computation, cognition, and communication - enables technological advances.
- It aims to develop and apply firm theoretical and mathematical foundations for the features that are common to all computational systems.
- Informatics has many aspects, and encompasses a number of existing academic disciplines - Artificial Intelligence, Cognitive Science and Computer Science.
- Thus Informatics provides a link between disciplines with their own methodologies and perspectives, bringing together a common scientific paradigm, common engineering methods and a pervasive stimulus from technological development and practical application.
- In our informatics system it is customized to predicting the critical temperature of superconductors.
- In this study, we take an entirely data-driven approach to create a statistical model that predicts.

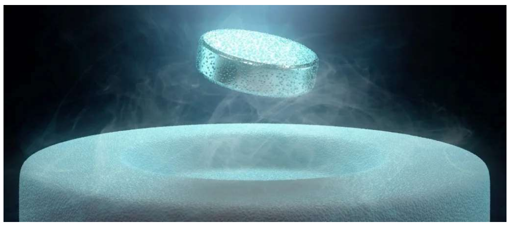
UPES
UNIVERSITY WITH A PURPOSE

INTRODUCTION

What is Superconductivity?

Superconductivity is a phenomenon whereby a charge moves through a material without resistance.

In other words Superconductivity, is a complete disappearance of electrical resistance in various solids when they are cooled below a characteristic temperature. This temperature, called the transition temperature, varies for different materials but generally is below 20 K (−253 °C).



UPES
UNIVERSITY WITH A PURPOSE

What is Critical Temp?

The **critical temperature** for **superconductors** is the **temperature** at which the electrical resistivity of a metal drops to zero. The transition is so sudden and complete that it appears to be a transition to a different phase of matter; this **superconducting** phase is described by the BCS theory.

How can we predict critical temp of superconductor?

The conventional way to predict the critical temp is to apply machine learning algorithms in it .There are numerous ways of creating database and apply by using different ML methods.

But we will Use CNN(Convolutional Neural Networks) to identify the critical temp by using minimum errors. CNN is a very good method because it involves thousand of variables in a single algorithm and give the approximate precise value.

UPES
UNIVERSITY WITH A PURPOSE

LITERATURE REVIEW

- In addition, machine learning has been proposed as a means to predict the material properties and selection from the materials database.
- This particular workflow calls for the materials database to be constructed using supercomputers and first principle calculations.
- This database is then trained through machine learning with the intention of using the trained machine to predict desirable materials and conduct targeted materials synthesis in experiments in a more efficient manner.

UPES
UNIVERSITY WITH A PURPOSE

LITERATURE REVIEW

Author	Data analysis Method	Informatics system Purpose
Kam Hamidieh.	• The Multiple Regression Model. • The XG-Boost Model.	A Data-Driven Statistical Model for Predicting the Critical Temperature of a Superconductor.
Shaobo Li 1, Yabo Dan 1,*, Xiang Li 1, Tiantian Hu 2, Rongzhi Dong 1, Zhuo Cao 1 and Jianjun Hu.	• A hybrid neural network (HNN) that combines a convolutional neural network (CNN) and long short-term memory neural network (LSTM).	Critical Temperature Prediction of Superconductors Based on Atomic Vectors and Deep Learning.
J.M. Rickman1,2, H.M. Chan2, M.P. Harmer2, J.A. Smeltzer2, C.J. Marvel2, A. Roy3 & G. Balasubramanian3.	• A multiple regression analysis and its generalization. • A canonical-correlation analysis (CCA).	The screening of multi-principal elements and high-entropy alloys.
Zhong-Li Liu,1,2,a) Peng Kang,2 Yu Zhu,3 Lei Liu,3 and Hong Guo.	• Multi-algorithm cross-validation. • Multi-step learning approach.	Layered high-TC of superconductors

UPES
UNIVERSITY WITH A PURPOSE

RESEARCH GAP

Using data-driven techniques, herein we propose a new strategy to extract relations from experimental databases with small data sizes.

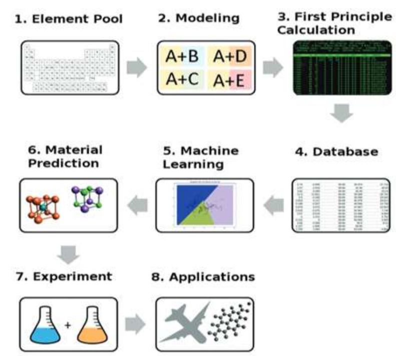
- First, we train a machine learning prediction model for the mechanical properties when process parameters and compositions of elements are inputted. Here, the prediction model with the highest prediction performance is selected from some supervised machine learning models.
- Second, the distributions of the process parameters and element compositions based on the prediction model are investigated to obtain the desired mechanical properties. Our strategy employs generator of new data by the prediction model.
- Third, we extract factors for the desired mechanical properties, including the influence degree at a glance, to promote the understanding of alloys.

UPES
UNIVERSITY WITH A PURPOSE

Methodology

Methodology to be followed to achieve the defined objectives is given below

- **Literature Analysis:** After studying the research papers we will be able to create an extensive alloy database from existing literature and can modify it in our project by analyzing them.
- **Database Preparation** : The database is the most essential part of any informatics system.
- **Data Analysis:** Now we will analyze the database and extract the information required in our project and this data can be implemented in Python.
- **Python Programming:** Now the main task is to learn the Python programming from the data we have extracted from papers and implement machine learning algorithms for structure and property prediction.
- **Validation of Model:** The final task is to validate our model and compare it with previous experiments and results.



UPES
UNIVERSITY WITH A PURPOSE

Challenges

1. Formulating appropriate database.
2. Data analysis from database.
3. Deriving co-relations between parameters.
4. Developing expertise in Python Programming for the informatics system as we are from mechanical engineering.
5. Implementation of machine learning algorithms.

UPES
UNIVERSITY WITH A PURPOSE

OBJECTIVE

1. Create an extensive database from existing literature and databases.
2. Implementing machine learning algorithms for structure and property prediction.
3. Model validation through comparison with reported/experimental results

UPES
UNIVERSITY WITH A PURPOSE

EXPECTED OUTCOMES

A machine learning model that can be used for:

- Predictive modeling of materials
- Identifying compositions that may lead to better/unique properties
- Understanding effect of different alloying additions

UPES
UNIVERSITY WITH A PURPOSE

COST ANALYSIS

<u>Direct costs:</u> No direct expenses in the current work plan. Based on results obtained, if experimental validation is required then characterization costs would be incurred (to be covered by mentor).

<u>Indirect costs :</u> Computational resources: Personal laptop/workstation, Google cloud servers (access will be provided by faculty), UPES high computation facility (if required)

TIME CHART

	SEP	OCT	NOV	DEC	JAN	FEB	MAR
LITERATURE REVIEW /LEARNING PYTHON							
DATABASE CREATION/LEARINING PYTHON							
DATA ANALYSIS / LEARINING PYTHON							
DERIVING INTER RELATION DEPENDENCIES.							
DEVELOPING INFORMATICS SYSTEM.							
MODEL VALIDATION							

UPES
UNIVERSITY WITH A PURPOSE

WORK DONE TILL DATE

DATABASE CREATION : Materials informatics is strongly dependent on large collections of data called big data. However, database management including collecting and organizing data can be quite problematic. A core part of materials informatics is the development of the database.here It is done in two parts

1. There are many databases being developed which vary in accessibility, themes, and depth. The superconductor data comes from the Superconducting Material Database maintained by Japan's National Institute for Materials Science (NIMS) at http: //supercon.nims.go.jp/index_en.html. Database is supported by the NIMS, a public institution based in Japan. The database contains a large list of superconductors, their critical temperatures, and the source references mostly from journal articles.
2. Element Data Preparation The element data are obtained by using the ElementData function from Mathematica .The first ionization energy data came from http://www.ptable.com/ and is merged with the Mathematical data.

UPES
UNIVERSITY WITH A PURPOSE

DATABASE

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	number_c	mean_ato	wtd_mea	gmean_at	wtd_gme	entropy_a	wtd_entr	range_ato	wtd_rang	std_atomi	wtd_std_a	mean_fie	wtd_mea	gmean_fi	wtd_gme	entropy_f	wtd_entr	range_fie	wtd_rang	std_fie	wtd_s
2	4	88.94447	57.86269	66.36159	36.11661	1.181795	1.062396	122.9061	31.79492	51.96883	53.62253	775.425	1010.269	718.1529	938.0168	1.305967	0.791488	810.6	735.9857	323.8118	355.
3	5	92.72921	58.51842	73.13279	36.3966	1.449309	1.057755	122.9061	36.16194	47.09463	53.97987	766.44	1010.613	720.6055	938.7454	1.544145	0.807078	810.6	743.1643	290.183	354.9
4	4	88.94447	57.88524	66.36159	36.12251	1.181795	0.97598	122.9061	35.7411	51.96883	53.65627	775.425	1010.82	718.1529	939.009	1.305967	0.77362	810.6	743.1643	323.8118	354.8
5	4	88.94447	57.87397	66.36159	36.11956	1.181795	1.022291	122.9061	33.76801	51.96883	53.6394	775.425	1010.544	718.1529	938.5128	1.305967	0.783207	810.6	739.575	323.8118	355.1
6	4	88.94447	57.84014	66.36159	36.11072	1.181795	1.129224	122.9061	27.84874	51.96883	53.58877	775.425	1009.717	718.1529	937.0256	1.305967	0.80523	810.6	728.8071	323.8118	356.3
7	4	88.94447	57.79504	66.36159	36.09893	1.181795	1.225203	122.9061	20.68746	51.96883	53.52115	775.425	1008.614	718.1529	935.0463	1.305967	0.824743	810.6	714.45	323.8118	357.8
8	4	88.94447	57.6823	66.36159	36.06947	1.181795	1.316857	122.9061	10.76564	51.96883	53.35156	775.425	1005.857	718.1529	930.1164	1.305967	0.841872	810.6	678.5571	323.8118	361.5
9	4	76.51772	57.17514	59.3101	35.89137	1.197273	0.94356	122.9061	36.4512	44.28946	52.92414	787.05	1011.484	734.2196	940.197	1.313008	0.776332	772	742.5	314.506	353.8
10	4	76.51772	56.80882	59.3101	35.77343	1.197273	0.98188	122.9061	34.83316	44.28946	52.53321	787.05	1011.541	734.2196	940.2943	1.313008	0.786865	772	738.5786	314.506	353.8
11	4	76.51772	56.44249	59.3101	35.65588	1.197273	1.016495	122.9061	33.21512	44.28946	52.13677	787.05	1011.597	734.2196	940.3917	1.313008	0.795977	772	734.6571	314.506	353.7
12	4	76.51772	55.70984	59.3101	35.42195	1.197273	1.077783	122.9061	29.97904	44.28946	51.32687	787.05	1011.71	734.2196	940.5864	1.313008	0.811136	772	726.8143	314.506	353.5
13	5	111.2736	63.71346	82.79332	37.93423	1.409442	1.335472	184.5906	27.84874	64.459	60.9031	821.54	1020.903	768.2333	949.1652	1.542797	0.896266	810.6	728.8071	303.9566	351.9
14	5	92.72921	58.20183	73.13279	36.2593	1.449309	1.026457	122.9061	36.93243	47.09463	53.81924	766.44	1010.716	720.6055	938.8772	1.544145	0.794064	810.6	745.125	290.183	354.8
15	5	92.72921	58.51842	73.13279	36.3966	1.449309	1.057755	122.9061	36.16194	47.09463	53.97987	766.44	1010.613	720.6055	938.7454	1.544145	0.807078	810.6	743.1643	290.183	354.9
16	5	92.72921	59.46818	73.13279	36.81165	1.449309	1.114758	122.9061	35.7411	47.09463	54.44786	766.44	1010.302	720.6055	938.3501	1.544145	0.831386	810.6	743.1643	290.183	355.2
17	5	92.72921	61.05111	73.13279	37.51393	1.449309	1.146919	122.9061	35.7411	47.09463	55.18273	766.44	1009.784	720.6055	937.6917	1.544145	0.84444	810.6	743.1643	290.183	355.5
18	5	69.17125	47.50532	54.87277	33.31907	1.419173	1.428952	121.3276	14.30396	41.80901	40.72776	753.08	1006.965	708.3233	943.3288	1.543038	0.943577	810.6	684.3846	290.5942	337.8
19	4	88.94447	57.87397	66.36159	36.11956	1.181795	1.022291	122.9061	33.76801	51.96883	53.6394	775.425	1010.544	718.1529	938.5128	1.305967	0.783207	810.6	739.575	323.8118	355.1
20	4	76.51772	56.80882	59.3101	35.77343	1.197273	0.98188	122.9061	34.83316	44.28946	52.53321	787.05	1011.541	734.2196	940.2943	1.313008	0.786865	772	738.5786	314.506	353.8
21	4	76.51772	57.54147	59.3101	36.00969	1.197273	0.899625	122.9061	38.06924	44.28946	53.30969	787.05	1011.428	734.2196	940.0997	1.313008	0.763668	772	746.4214	314.506	353.9
22	4	76.51772	57.17514	59.3101	35.89137	1.197273	0.94356	122.9061	36.4512	44.28946	52.92414	787.05	1011.484	734.2196	940.197	1.313008	0.776332	772	742.5	314.506	353.8
23	4	76.51772	56.25933	59.3101	35.59726	1.197273	1.032709	122.9061	32.4061	44.28946	51.93645	787.05	1011.625	734.2196	940.4404	1.313008	0.800111	772	732.6964	314.506	353.7
24	4	76.51772	56.07617	59.3101	35.53872	1.197273	1.048294	122.9061	31.59708	44.28946	51.7347	787.05	1011.654	734.2196	940.4891	1.313008	0.804002	772	730.7357	314.506	353.

db_superconductors Sheet1

Ready Average: 534.3507452 Count: 1743648 Sum: 931675791.4 100%

Size of database-Excel denotation
Count-1743648
Sum-931675791.4

UPES
UNIVERSITY WITH A PURPOSE

DATABASE

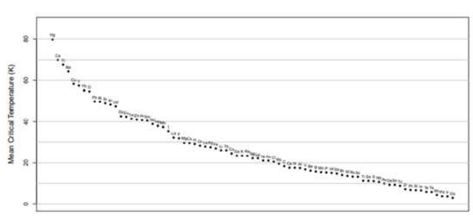
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	number_c	mean_ato	wtd_mear	gmean_at	wtd_gmea	entropy_a	wtd_entrc	range_ato	wtd_range	std_atomi	wtd_std_a	mean_fie	wtd_mear	gmean_fi	wtd_gmea	entropy_f	wtd_entrc	range_fie	wtd_range	std_fie	wtd_s
2	4	88.94447	57.86269	66.36159	36.11661	1.181795	1.062396	122.9061	31.79492	51.96883	53.62253	775.425	1010.269	718.1529	938.0168	1.305967	0.791488	810.6	735.9857	323.8118	355.
3	5	92.72921	58.51842	73.13279	36.3966	1.449309	1.057755	122.9061	36.16194	47.09463	53.97987	766.44	1010.613	720.6055	938.7454	1.544145	0.807078	810.6	743.1643	290.183	354.9
4	4	88.94447	57.88524	66.36159	36.12251	1.181795	0.97598	122.9061	35.7411	51.96883	53.65627	775.425	1010.82	718.1529	939.009	1.305967	0.77362	810.6	743.1643	323.8118	354.8
5	4	88.94447	57.87397	66.36159	36.11956	1.181795	1.022291	122.9061	33.76801	51.96883	53.6394	775.425	1010.544	718.1529	938.5128	1.305967	0.783207	810.6	739.575	323.8118	355.1
6	4	88.94447	57.84014	66.36159	36.11072	1.181795	1.129224	122.9061	27.84874	51.96883	53.58877	775.425	1009.717	718.1529	937.0256	1.305967	0.80523	810.6	728.8071	323.8118	356.3
7	4	88.94447	57.79504	66.36159	36.09893	1.181795	1.225203	122.9061	20.68746	51.96883	53.52115	775.425	1008.614	718.1529	935.0463	1.305967	0.824743	810.6	714.45	323.8118	357.8
8	4	88.94447	57.6823	66.36159	36.06947	1.181795	1.316857	122.9061	10.76564	51.96883	53.35156	775.425	1005.857	718.1529	930.1164	1.305967	0.841872	810.6	678.5571	323.8118	361.5
9	4	76.51772	57.17514	59.3101	35.89137	1.197273	0.94356	122.9061	36.4512	44.28946	52.92414	787.05	1011.484	734.2196	940.197	1.313008	0.776332	772	742.5	314.506	353.8
10	4	76.51772	56.80882	59.3101	35.77343	1.197273	0.98188	122.9061	34.83316	44.28946	52.53321	787.05	1011.541	734.2196	940.2943	1.313008	0.786865	772	738.5786	314.506	353.8
11	4	76.51772	56.44249	59.3101	35.65588	1.197273	1.016495	122.9061	33.21512	44.28946	52.13677	787.05	1011.597	734.2196	940.3917	1.313008	0.795977	772	734.6571	314.506	353.7
12	4	76.51772	55.70984	59.3101	35.42195	1.197273	1.077783	122.9061	29.97904	44.28946	51.32687	787.05	1011.71	734.2196	940.5864	1.313008	0.811136	772	726.8143	314.506	353.5
13	5	111.2736	63.71346	82.79332	37.93423	1.409442	1.335472	184.5906	27.84874	64.459	60.9031	821.54	1020.903	768.2333	949.1652	1.542797	0.896266	810.6	728.8071	303.9566	351.9
14	5	92.72921	58.20183	73.13279	36.2593	1.449309	1.026457	122.9061	36.93243	47.09463	53.81924	766.44	1010.716	720.6055	938.8772	1.544145	0.794064	810.6	745.125	290.183	354.8
15	5	92.72921	58.51842	73.13279	36.3966	1.449309	1.057755	122.9061	36.16194	47.09463	53.97987	766.44	1010.613	720.6055	938.7454	1.544145	0.807078	810.6	743.1643	290.183	354.9
16	5	92.72921	59.46818	73.13279	36.81165	1.449309	1.114758	122.9061	35.7411	47.09463	54.44786	766.44	1010.302	720.6055	938.3501	1.544145	0.831386	810.6	743.1643	290.183	355.2
17	5	92.72921	61.05111	73.13279	37.51393	1.449309	1.146919	122.9061	35.7411	47.09463	55.18273	766.44	1009.784	720.6055	937.6917	1.544145	0.84444	810.6	743.1643	290.183	355.5
18	5	69.17125	47.50532	54.87277	33.31907	1.419173	1.428952	121.3276	14.30396	41.80901	40.72776	753.08	1006.965	708.3233	943.3288	1.543038	0.943577	810.6	684.3846	290.5942	337.8
19	4	88.94447	57.87397	66.36159	36.11956	1.181795	1.022291	122.9061	33.76801	51.96883	53.6394	775.425	1010.544	718.1529	938.5128	1.305967	0.783207	810.6	739.575	323.8118	355.1
20	4	76.51772	56.80882	59.3101	35.77343	1.197273	0.98188	122.9061	34.83316	44.28946	52.53321	787.05	1011.541	734.2196	940.2943	1.313008	0.786865	772	738.5786	314.506	353.8
21	4	76.51772	57.54147	59.3101	36.00969	1.197273	0.899625	122.9061	38.06924	44.28946	53.30969	787.05	1011.428	734.2196	940.0997	1.313008	0.763668	772	746.4214	314.506	353.9
22	4	76.51772	57.17514	59.3101	35.89137	1.197273	0.94356	122.9061	36.4512	44.28946	52.92414	787.05	1011.484	734.2196	940.197	1.313008	0.776332	772	742.5	314.506	353.8
23	4	76.51772	56.25933	59.3101	35.59726	1.197273	1.032709	122.9061	32.4061	44.28946	51.93645	787.05	1011.625	734.2196	940.4404	1.313008	0.800111	772	732.6964	314.506	353.7
24	4	76.51772	56.07617	59.3101	35.53872	1.197273	1.048294	122.9061	31.59708	44.28946	51.7347	787.05	1011.654	734.2196	940.4891	1.313008	0.804002	772	730.7357	314.506	353.

db_superconductors

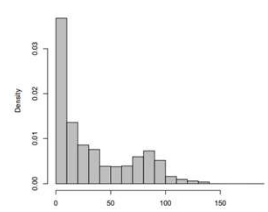
UPES
UNIVERSITY WITH A PURPOSE

DATA ANALYSIS

MODEL ANALYSIS : We have just started to investigate the database that we have. In this section we will be discussing the results of the Linear, poly regression model. Neural network tends to be the most accurate one therefore will be performing that for the most accurate system. For the start We tried plotting few graphs models to visualize the data dependencies.



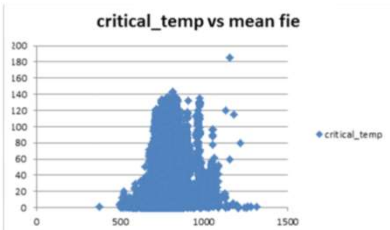
This figure shows the mean superconducting critical temperature grouped by elements. On average, mercury containing materials had the highest superconducting critical temperature followed by calcium and so on.

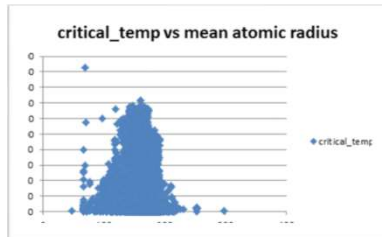


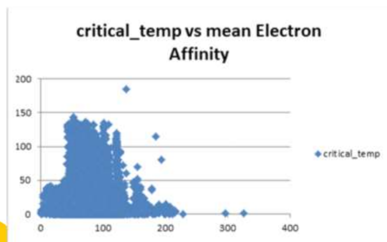
This figure shows the distribution of the superconducting critical temperatures (K) of all 21,263 superconductors.

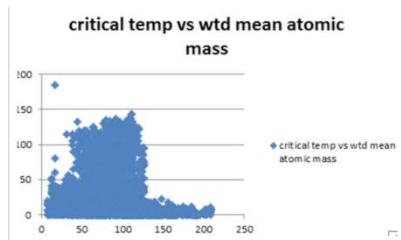
UPES
UNIVERSITY WITH A PURPOSE

Graphs plotted : For statistical understanding of Data.
This not all the data plots .

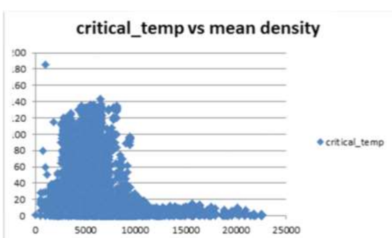


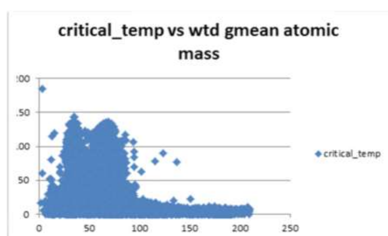


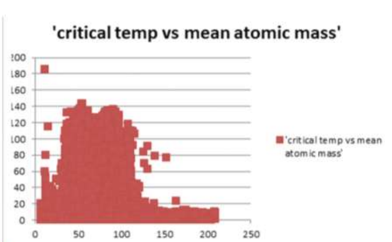


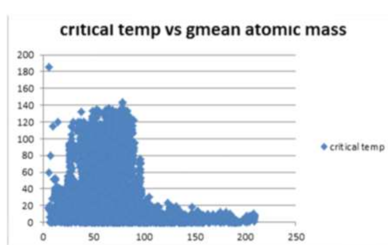


UPES
UNIVERSITY WITH A PURPOSE









UPES
UNIVERSITY WITH A PURPOSE

Data set creation : After visual analysis of variation of critical temperatures along with different variable we narrowed down to 8 variable from 80 features are derived

Variable	Units	Description
Atomic mass	atomic mass units (AMU)	total proton and neutron rest masses
Frist ionization energy	Joules per mole (kJ/mol)	energy required to remove a valence electron
Atomic radius	picometer (pm)	calculated atomic radius
Density	kilograms per meters cubed (kg/m3)	density at standard temperature and pressure
Electron affinity	kilo-Joules per mole (kJ/mol)	energy required to add an electron to a neutral atom
Fusion heat	kilo-Joules per mole (kJ/mol)	energy to change from solid to liquid without temperature change
Thermal conductivity	watts per meter-Kelvin (W/(m × K))	Thermal conductivity co-efficient
Valence	No unit	Typical number of chemical bonds formed by the element.

Above given is table of 8 elemental features which are used for creating alloy features from every element present in periodic table .

Since we are considering for alloy we can convert elemental property to some kind of alloy property .

For each of these 8 elemental properties we have extended each elemental properties into 10 alloy properties.

Feature and description	Formula	Sample values
Mean	$= \mu = (t1 + t2)/2$	35.5
Weighted mean	$= \nu = (p1t1) + (p2t2)$	44.43
Geometric mean	$= (t1t2)\ 1/2$	33.23
Weighted geometric mean	$= (t1)\ p1\ (t2)\ p2$	43.21
Entropy	$= -w1\ \ln(w1) - w2\ \ln(w2)$	0.63
Weighted entropy	$= -A\ \ln(A) - B\ \ln(B)\ 0.26$	0.26
Range	$= t1 - t2\ (t1 > t2)$	25
Weighted range	$= p1t1 - p2t2$	37.86
Standard deviation	$= [(1/2)((t1 - \mu)\ 2 + (t2 - \mu)\ 2\ )]1/2$	12.5
Weighted standard deviation	$= [p1(t1 - \nu)\ 2 + p2(t2 - \nu)\ 2\ )]1/2$	8.75

This table summarizes the procedure for feature extraction from material's chemical formula.

Feature extraction :

The feature extraction process through a detailed example: Consider Re7Zr1 with Tc = 6.7 K, focusing on the features extracted based on thermal conductivity. Rhenium and Zirconium's thermal conductivity coefficients are t1 = 48 and t2 = 23 W/(m×K) respectively.

- The ratios of the elements in the material are used to define features:

: p1 = 6 6 + 1 = 6 7 , p2 = 1 6 + 1 = 1 7 .

- The fractions of total thermal conductivities are used as well:

: w1 = t1 t1 + t2 = 48 48 + 23 = 48 71 , w2 = t2 t1 + t2 = 23 48 + 23 = 23 71 .

We need a couple of intermediate values based on equations eq(1) and eqn(2):

A = p1w1 p1w1 + p2w2 ≈ 0.926

B = p2w2 p1w1 + p2w2 ≈ 0.074.

Feature extraction :

Once we have obtained the values p1, p2, w1, w2, A, and B, we can extract 10 features from Rhenium and Zirconium's thermal conductivities . We repeat the same process above with the 8 variables

For example, for features based on atomic mass, just replace t1 and t2 with the atomic masses of Rhenium and Zirconium respectively, then carry on with the calculations of p1, p2, w1, w2, A, B, and finally calculate the 10 features defined in table (2). This gives us 8 × 10 = 80 features. One additional features, a numeric variable counting the number of elements in the supercondutor, is also extracted. We end up with 81 features in total..

Now with the help of 80 alloy features and elemental composition table the training set is narrowed down .which is converted into CSV (comma separated file
) which can be directly feed into the program .

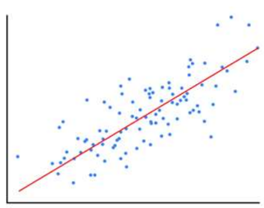
Regression Modelling

Regression modelling:
Regression modelling is a form of predictive modelling techniques which investigates the relationship between dependent variables(target) and independent variables(parameters). This technique is used for forecasting ,time series modelling and finding the causal effect relationship between the variables.
There are various kind of regression techniques available to make predictions, mostly these are driven by three metric -
1.Number of independent variables.
2.Shape of the regression line.
3. Type of dependent variables.

Linear regression:

One of the most widely used modelling techniques . In this techniques the dependent variable is continuous ,independent can be continuous or discrete whereas the nature of regression line is *straight line.*

. Graph example of linear regression.

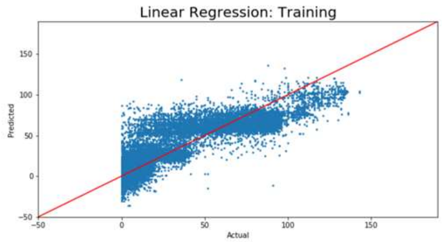


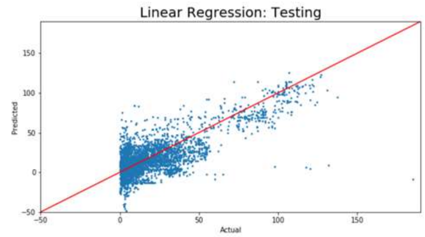
Y=f(x1,x2,x3,………….,x80).

Y= critical temperature , X1,x2,x3: alloy features.

Linear Regression establishes a relationship between dependent variable (Y) and one or more independent variables (X) using a best fit straight line (also known as regression line).
Assumption : critical temperature is function of 80 feature .
The training dataset include 80 alloy feature along with the elemental data composition table on which the the linear model is trained , which was further divide into ratio of 80:20 as training set and testing sets.
Platform used for liner regression coding is Python

Parity plot : Linear regression



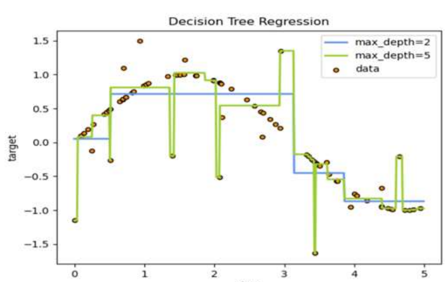


R2 score for training set obtained .73 which is decent but when testing set was feed into the system R2 score decreased to .49 which is not satisfactory to predict the value .
After looking at the graph we had a rough idea that the result were not close to accurate.
Percentage error became quite high for testing sets.at high temperature (critical) results were close to the actual experimental values but with low temperature the %error high which is not acceptable. Not a good fit was obtained , at high temperature linear model gave moderate value but at low temperature % error became high . therefore the predicted values were too off range that it could be considered on.

UPES
UNIVERSITY WITH A PURPOSE

Decision Trees (DTs)

It is a non-parametric supervised learning method used for classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. A tree can be seen as a piecewise constant approximation.Requires little data preparation. Other techniques often require data normalisation, dummy variables need to be created and blank values to be removed. Note however that this module does not support missing values . The cost of using the tree (i.e., predicting data) is logarithmic in the number of data points used to train the tree.

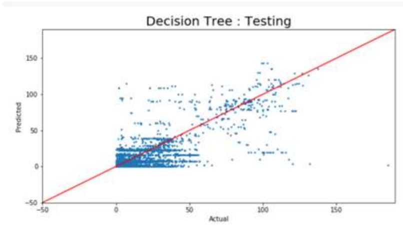


Since the linear model results were not very satisfactory to identify and the variable's effect on the critical temperature decision tree is now modelled .Same dataset of 80 alloy elemental database is feed to the system, coding performed on python

Thermal Energy Storage (TES) System for Concentrated Solar Power (CSP) Application-Synopsis

UPES
UNIVERSITY WITH A PURPOSE

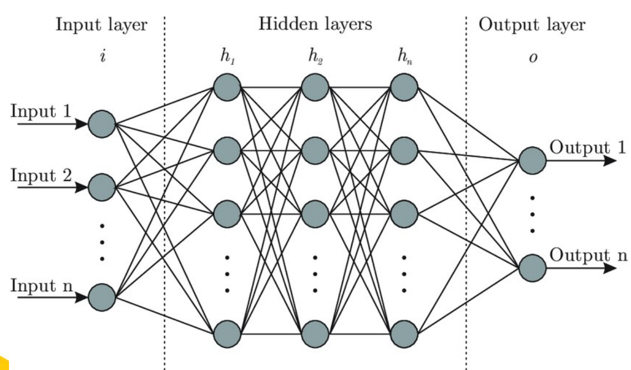
Parity plot : Decision tree regression

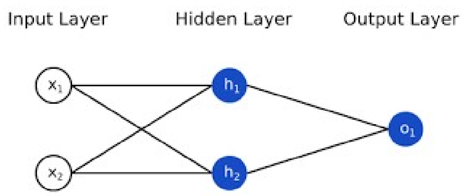


Score for training set =98.37 %
Score for testing set = 43.60 %
For training set the results were more precise than linear , ac considerable improvement was seen which can be taken into consideration ,but when the model was fed with testing set the score dropped drastically ,which is better than linear but a significant growth which can be considered.
Over-fit observed : the model 's result for training set had much better accuracy for training test but for testing set the results were poor . For training decision tree model fits well but for the testing is does'nt fit well. The model can not make a significant good prediction for unseen experimental data.
Clustering of prediction observed: Many points had same predicted values in testing model results whereas the actual value is very different which can be observed by the horizontal lines in parity plot for testing model.

Neural Network

A **neural network** is a series of algorithms that endeavors to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates.
neural networks refer to systems of neurons, either organic or artificial in nature.





Connected to every layer of hidden units.
Can add more hidden units and each will contribute to other than other.
This will decide what output to send , it is biased by one value and then it is taken by some function.
generally SIMPLOID FUNCTION.

CONCLUSION

After plotting the result for both regression model (i:e linear and decision) , it can be found that the decision tree has better performance than linear model but not enough , such that predicted value can be considered as the training set seem to have high value of accuracy but not testing sets.

The results improved after applying decision tree but not to such extent that it can have a significant considerations.

Future work:

Looking forward to apply random forest method such that clustering of data and over-fit can be eliminated from decision tree such that better inter-dependencies can be derived.

Followed up by neural network such that the model may become capable to predict the superconductor 's critical temperature with the simple input of composition into the model

<u>Sept. 2020 – Dec 2020:</u> Literature review, Creation of database, Data Analysis, Learning Python programming.

<u>Jan 2021 – March 2021:</u> Deriving inter-relation Co-dependencies, Developing informatics system (Implementation of machine learning algorithms : neural networks/decision trees/support vector machines), Model validation.

REFERNCES

[1]. Kam HamidiehUniversity of Pennsylvania, Wharton, Statistics DepartmentScientific Reports 3, 1–6 (2013).
[2]. Shaobo Li 1, Yabo Dan 1,*, Xiang Li 1, Tiantian Hu 2, Rongzhi Dong 1, Zhuo Cao 1 and Jianjun Hu 1,3,*Published: 8 February 2020
[3]. Materials informatics for the screening of multi-principal elements and high-entropy alloys J.M. Rickman1,2, H.M. Chan2, M.P. Harmer2, J.A. Smeltzer2, C.J. Marvel2, A. Roy3 & G. Balasubramanian3 (2019) 10:2618 |
[4]. Material informatics for layered high-TC Superconductors Zhong-Li Liu , Peng Kang , Yu Zhu, Lei Liu, and Hong Guo ,APL Mater. 8, 061104 (2020).