SRM INSTITUTE OF SCIENCE & TECHNOLOGY
Deemed to be University u/s 3 of UGC Act, 1956

PREDICTION OF PARKINSON'S DISEASE

21CSC305P - MACHINE LEARNING

REVIEW - I

BATCH 12 :-

P. DIMPUL DURGA - RA2211003011478

K. SAILAGA - RA2211003011482

T. RAGHU - RA2211003011483

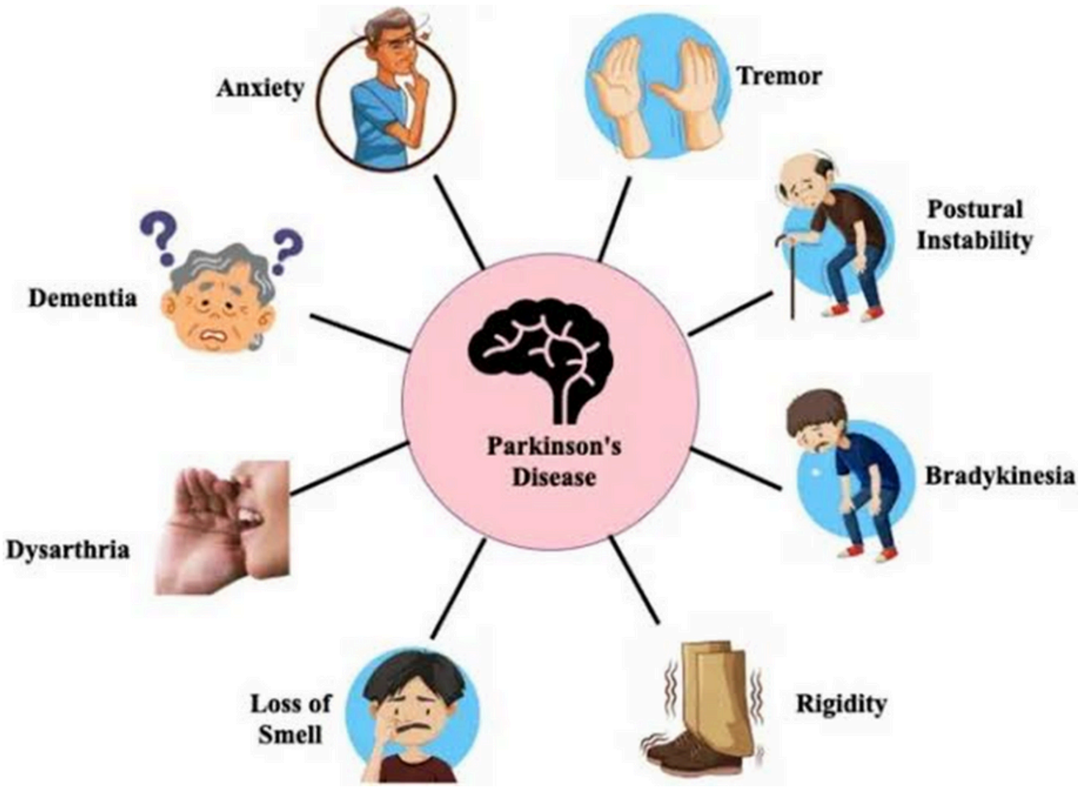
K. CHARMI AKSHITA - RA2211003011485

INTRODUCTION

- PARKINSON'S DISEASE IS A PROGRESSIVE NEUROLOGICAL DISORDER THAT PRIMARILY AFFECTS MOVEMENT.
- IT IS CAUSED BY THE DEGENERATION OF DOPAMINE-PRODUCING NEURONS IN A SPECIFIC AREA OF THE BRAIN CALLED THE SUBSTANTIA NIGRA.
- DOPAMINE IS A NEUROTRANSMITTER THAT PLAYS A KEY ROLE IN SENDING MESSAGES TO THE PART OF THE BRAIN THAT CONTROLS MOVEMENT AND COORDINATION.
- THE EXACT CAUSE OF PARKINSON'S DISEASE IS NOT KNOWN, BUT A COMBINATION OF GENETIC AND ENVIRONMENTAL FACTORS IS THOUGHT TO PLAY A ROLE. SOME CASES ARE LINKED TO SPECIFIC GENETIC MUTATIONS, WHILE ENVIRONMENTAL EXPOSURES TO TOXINS MAY INCREASE THE RISK.

PROBLEM STATEMENT

- PREDICTING THE PARKINSON'S DISEASE USING VARIOUS TESTS SUCH AS JITTER, SHIMER, HARMONAL TO NOISE RATIO(HNR), ETC... AND TRAINING THE MACHINE TO PREDICT THE OUTCOME THROUGH RANDOM FOREST ALGORITHM AND SVM'S.



DATA EXTRACTION

DATA IS EXTRACTED FROM THE VARIOUS SITES SUCH AS KAGGLE, ONLINE SITES AND YOU TUBE.

ID	Age	Sex	Jitter	Shimmer	HNR	RPDE	DFA	Spread1	Spread2	PPE	Parkinsons
1	40	M	0.01638609177641413	0.022630787802965846	0.19623967374499934	0.36164026934321675	0.2576330926461961	0.5805325958017569	0.4935324474203187	0.39358360142997617	0
2	43	M	0.014299967804707131	0.023335073956751825	0.11472476726719898	0.3495224557701921	0.2953923591046441	0.5790569700685467	0.4910381390725521	0.3500781716368703	0
3	43	F	0.010357426825004838	0.025723170159586254	0.14424158153889982	0.3230476070159513	0.2247145165991135	0.5885645250674025	0.44381927803188037	0.31112029245745315	1
4	79	F	0.01770128124191582	0.028948697402009824	0.14940925279881825	0.35351480485798903	0.2961632975675131	0.5050741830639421	0.4492400695373781	0.3475181934284344	0
5	49	F	0.01502105580724113	0.021762816401400147	0.18602812365183047	0.3065954659700089	0.2524026094590923	0.5982216859910011	0.4062090597485206	0.33601307939534797	0
6	59	F	0.017861884994650606	0.022796787975992203	0.16279172431961692	0.3553714157590274	0.2728386962090005	0.5426461491788597	0.4839372866167314	0.3034104189980439	1
7	61	F	0.017480227993432698	0.02581679843147862	0.17745664242016157	0.3854465441873847	0.2781340896758797	0.5192982107945779	0.4780487537972791	0.3349709797619094	1
8	76	M	0.01793567368048352	0.024543342421615152	0.1485096845431291	0.31409724125629007	0.21646092288967353	0.5970022783367631	0.43932889451979223	0.3976340817792969	1
9	63	F	0.013006511586967669	0.024473228872557252	0.16484137022418283	0.3325352186740872	0.29747446050002135	0.5825182322876131	0.447708726444492	0.32596493653841313	0
10	46	M	0.01800798599067631	0.028207342620964065	0.17354093600258585	0.36311232710518	0.2976120577808166	0.5929723552383818	0.46593617967637135	0.3724002770812333	0
11	64	M	0.015488463284661674	0.02923878302511361	0.17378812839173013	0.33593012786871296	0.28661700552572067	0.5523153725753763	0.4465579036813512	0.315413984555746	1
12	64	M	0.014733262004491432	0.02481306966069882	0.10017378267846724	0.3415003027147682	0.28433548154726485	0.5747741849615413	0.427319039405026	0.31693204844998507	1
13	52	F	0.016751259138207568	0.026873517954472737	0.10999080737300607	0.333845089816002	0.23262644217301998	0.5519464037774283	0.4297485443068282	0.32964909701605216	0
14	41	F	0.010213586829383484	0.02801058711027394	0.13636604411044162	0.3482669895603961	0.2852066460701215	0.5539450113811455	0.45173130080204527	0.3014614720515555	1
15	78	M	0.011023168158875942	0.02518366410387164	0.10611974471035418	0.3373195953727305	0.2888327957126259	0.5951500964407036	0.49945310307798507	0.34836453641618803	0
16	79	M	0.01292177365167522	0.022943164089537352	0.1382989702367717	0.3144248323245875	0.226652645122604	0.5738356983642098	0.44899256787427066	0.3969825503731062	1
17	63	M	0.019829901099578398	0.026380845977222837	0.13895668484721924	0.3198812796198674	0.21706562199270088	0.5566777806847493	0.40557974199464397	0.32815309114449154	1
18	64	M	0.01139745778864955	0.025851091003060593	0.192023353727774	0.30792465900232624	0.2756978749143696	0.539273526443214	0.49386954162712937	0.3669827562064094	1
19	57	M	0.013305963007592103	0.0290156281973961	0.1567071854894691	0.34676570677191004	0.25777352938932435	0.5164327638614411	0.4305970896091819	0.3844281114411363	0
20	77	M	0.010510530646919236	0.02052407036769856	0.17063512805336584	0.38454039345087343	0.21078631998351594	0.5115730374575417	0.4039825376734112	0.39394560179124194	0
21	65	F	0.013312688802927262	0.029101313871334668	0.18871986820862285	0.3090551083385189	0.23995055990581043	0.5299666145536622	0.4072285761395784	0.3025326332204563	0
22	53	M	0.013203262865108317	0.025344320281653596	0.14421293206493174	0.36587254975151684	0.23319999451656265	0.5692804345197324	0.4171134604109579	0.3845118424611016	0
23	48	M	0.01946807170944522	0.02015676116049914	0.10019621445973823	0.34202231604535027	0.26606908139969726	0.5191552618765709	0.4634156407025436	0.3081399586270396	1
24	49	M	0.018451540869957206	0.02344702178767788	0.1285336024625414	0.38713810215947375	0.24204051540181176	0.5131462136590706	0.43679860824570244	0.33133866714029836	1
										0.37426521103683336	1

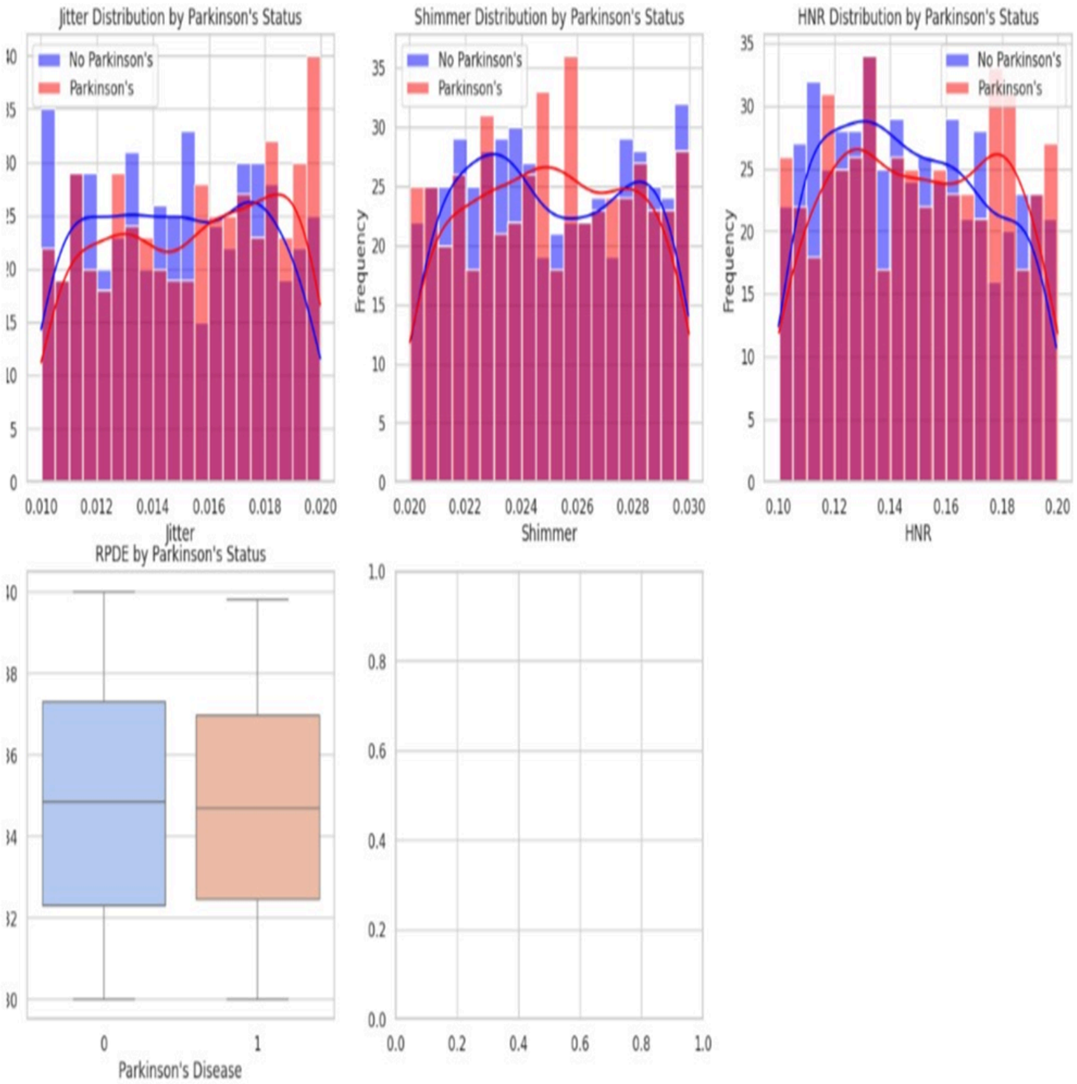
DATA EXPLORATION AND VISUALIZATION

DATA UNDERSTANDING

INITIAL DATA EXPLORATION REVEALS KEY CHARACTERISTICS OF THE DATASET, INCLUDING THE DISTRIBUTION OF FEATURES AND POTENTIAL RELATIONSHIPS BETWEEN THEM. WE EMPLOY DESCRIPTIVE STATISTICS AND VISUALIZATIONS TO GAIN INSIGHTS INTO THE DATA.

VISUAL DATA INSIGHTS

HISTOGRAMS, SCATTER PLOTS, AND BOX PLOTS ARE USED TO VISUALIZE THE DISTRIBUTIONS OF FEATURES AND UNCOVER POTENTIAL PATTERNS OR ANOMALIES. THESE VISUALIZATIONS HELP US UNDERSTAND THE NATURE OF THE DATA AND IDENTIFY POTENTIALLY IMPORTANT FEATURES.

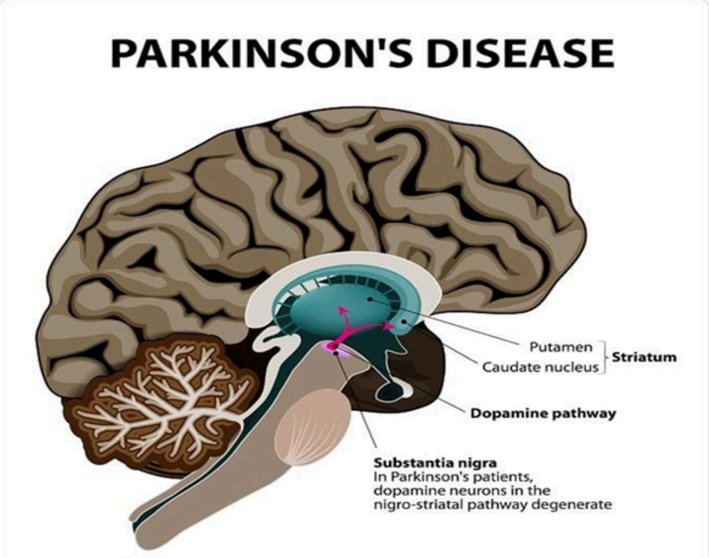


Pair Plot of Age, Jitter, and Shimmer

FEATURE ANALYSIS USING DATA PLOTS

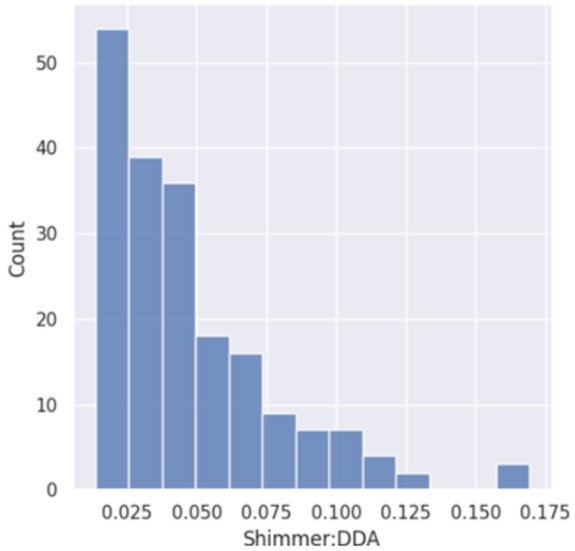
IDENTIFYING KEY FEATURES

SCATTER PLOTS AND CORRELATION MATRICES ARE CRUCIAL IN IDENTIFYING FEATURES THAT HAVE A STRONG CORRELATION WITH THE PRESENCE OF PARKINSON'S DISEASE. THIS HELPS US UNDERSTAND THE RELATIONSHIPS BETWEEN VARIABLES AND SELECT RELEVANT FEATURES FOR OUR MODEL.



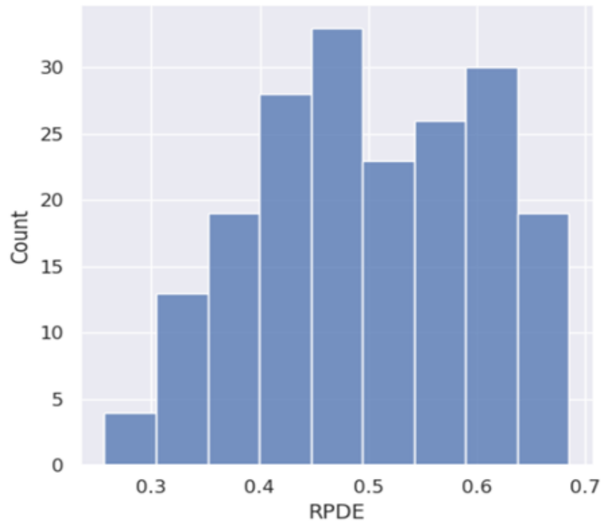
FEATURE IMPORTANCE

WE EXAMINE THE IMPACT OF VARIOUS FEATURES ON THE PREDICTION CONTRIBUTION OF EACH FEATURE TO THE MODEL'S PERFORMANCE. THIS HELPS US PRIORITIZE FEATURES OUTCOME BY ANALYZING THE AND UNDERSTAND THEIR SIGNIFICANCE IN PREDICTING PARKINSON'S DISEASE.



EXPLORING RELATIONSHIPS

SCATTER PLOTS AND BOX PLOTS ALLOW US TO OBSERVE THE DISTRIBUTION OF FEATURES WITHIN DIFFERENT CLASSES, REVEALING POTENTIAL DIFFERENCES BETWEEN INDIVIDUALS WITH AND WITHOUT PARKINSON'S DISEASE.



PANDAS, MATPLOTLIB AND SEABORN FOR DATA HANDLING AND VISUALIZATION

DATA MANIPULATION WITH PANDAS

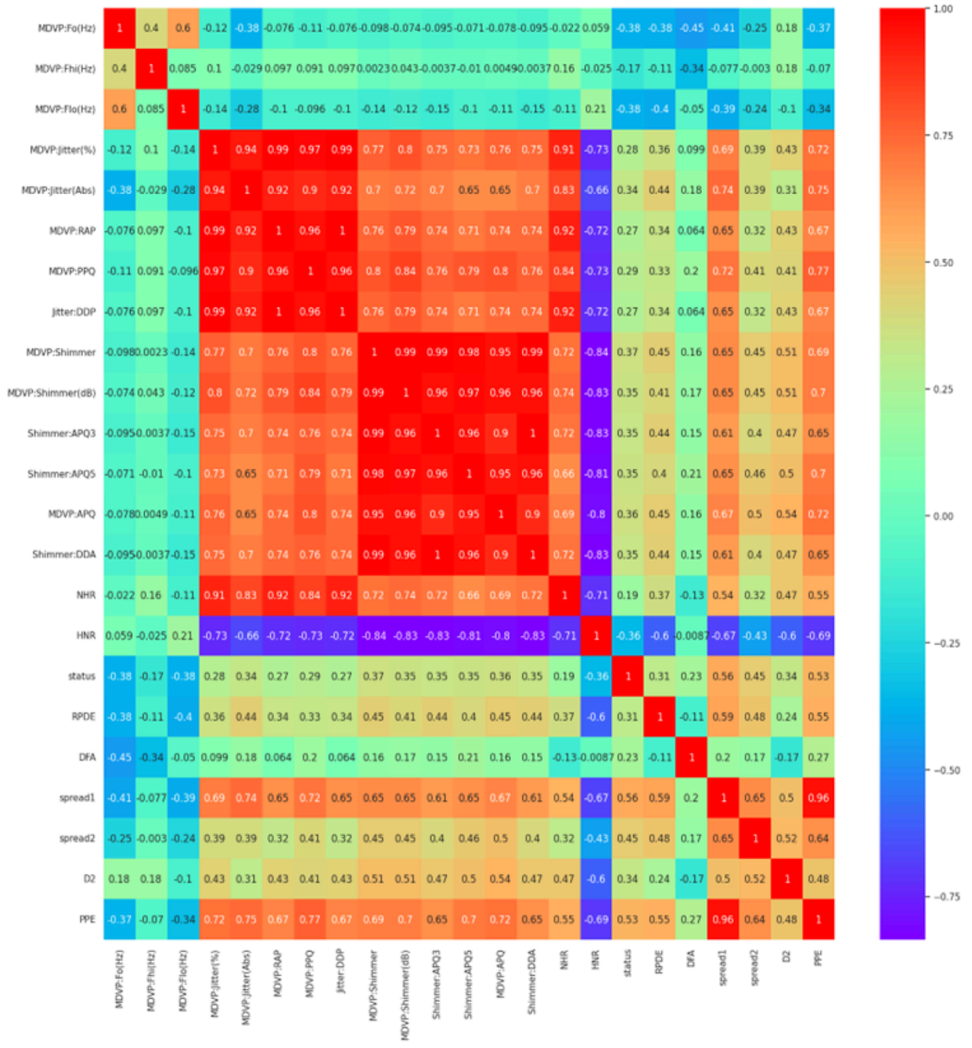
PANDAS IS USED TO LOAD, CLEAN, AND PREPARE THE DATA FOR ANALYSIS. WE LEVERAGE ITS POWERFUL DATA STRUCTURES AND FUNCTIONS TO EFFICIENTLY HANDLE THE DATASET AND PERFORM DATA TRANSFORMATIONS.

VISUALIZATIONS WITH MATPLOTLIB

MATPLOTLIB SERVES AS THE FOUNDATION FOR CREATING VARIOUS TYPES OF PLOTS. WE USE ITS EXTENSIVE LIBRARY OF FUNCTIONS TO GENERATE VISUALIZATIONS THAT REVEAL INSIGHTS AND TRENDS WITHIN THE DATA.

ENHANCING VISUALIZATIONS WITH SEABORN

SEABORN BUILDS UPON MATPLOTLIB, PROVIDING A HIGHER-LEVEL INTERFACE FOR CREATING AESTHETICALLY PLEASING AND INFORMATIVE STATISTICAL VISUALIZATIONS. IT SIMPLIFIES THE PROCESS OF GENERATING COMPLEX PLOTS AND ENHANCES THE VISUAL APPEAL OF OUR ANALYSES.



PREDICTIVE MODEL DEVELOPMENT

MODEL SELECTION

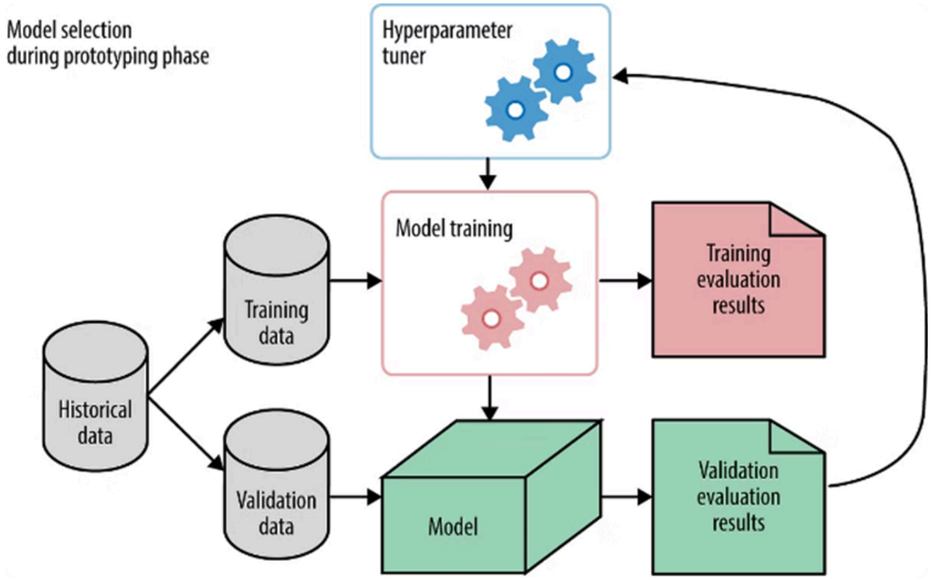
WE EVALUATE DIFFERENT MACHINE LEARNING ALGORITHMS LIKE LOGISTIC REGRESSION, DECISION TREES, AND SUPPORT VECTOR MACHINES, CONSIDERING THEIR STRENGTHS AND WEAKNESSES IN RELATION TO OUR SPECIFIC DATASET AND TASK.

HYPERPARAMETER TUNING

OPTIMIZATION OF MODEL PARAMETERS IS CRUCIAL FOR ACHIEVING THE BEST POSSIBLE PERFORMANCE. WE EXPLORE DIFFERENT HYPERPARAMETER COMBINATIONS TO FINE-TUNE THE MODEL AND FIND THE OPTIMAL SETTINGS FOR OUR DATASET.

MODEL EVALUATION

WE EMPLOY METRICS LIKE ACCURACY, PRECISION, RECALL, AND F1-SCORE TO EVALUATE THE MODEL'S PERFORMANCE. THIS ALLOWS US TO ASSESS THE MODEL'S ABILITY TO CORRECTLY CLASSIFY INDIVIDUALS WITH AND WITHOUT PARKINSON'S DISEASE.



FEATURE ENGINEERING AND SELECTION

Feature Engineering	Creating new features or transforming existing ones to improve model performance. This involves combining features, generating interactions, or applying domain-specific knowledge.
Feature Selection	Choosing the most relevant features for the model. This helps reduce noise and improve model accuracy by focusing on the most influential variables.