

INFORME TERCER
SPRINT CHEC



Análisis Estadístico Pandemia y no Pandemia:

Con el fin de saber que tanto cambió el comportamiento del sistema eléctrico de CHEC durante el periodo de la pandemia, se hizo necesario hacer un análisis estadístico de cada una de las barras de tensión y de los circuitos de red, tanto en época de pandemia como en época de no pandemia; para esto se procedió a calcular momentos estadísticos como la media y la desviación estándar para cada una de las variables de interés (VA, VB, VC, IA, IB, IC, P, Q) tanto en periodo de pandemia como en periodo de no pandemia; esto con el fin de posteriormente para cada circuito de red o barra de tensión calcular un porcentaje de cambio de los momentos estadísticos mencionados, obteniendo los siguientes resultados:

Circuito o Barra	Porcentaje de Cambio Media VA
MTO30	1542.1123256934156
BUM30	6.778433812291101
BQE30	0.9974814143453972
MAZ23	0.9656963836237676
VIR23	0.9567973105622168
IRR23	0.7022153255613618
HER40	0.6616672410020393
MNA30	0.5883369934225494
NRA30	0.54918009325566
HER46	0.4994902043722731
PUR46	0.4955826507836964
MNA23	0.4001465817996724
PSO10	0.367090211990049
FEL23	0.2332215995762482
MAN23	0.2267643248473707
QHI30	0.2126754941015002
DOR10	0.1891804140873274
AMR23	0.1831777269191204
GRO23	0.1796877334895148
PSO23	0.167340640616726
MTT23	0.1192288478872423
EDE30	0.1182990711154785
SIO30	0.0996193906703489
BOA23	0.098488628098455
MGT23	0.0983984353142599
VIR30	0.0954244421280599

Figura 1. Porcentaje de Cambio de la media de VA con respecto a la Pandemia.

Circuito o Barra	Porcentaje de Cambio Media VB
BUM30	6.755973819122812
BQE30	1.0
MAZ23	0.9656472446932414
VIR23	0.9578256484952076
HER46	0.8010883150843858
MAN10	0.7842364726118933
IRR23	0.70215941635752
HER40	0.6648142503250948
MNA30	0.5688854550841486
NRA30	0.5491660037313499
PUR46	0.4950202542582579
PSO10	0.4606963031697706
MNA23	0.3984174691132076
EDE30	0.3333646395913138
FEL23	0.2428013488230967
MAN23	0.2269320092110623
QHI30	0.2110561409009313
DOR10	0.1862258153215643
GRO23	0.179416080611936
AMR23	0.1664931242204871
PSO23	0.165231120944409
MLA23	0.1027197270848081
MTT23	0.1002042890287007
BOA23	0.0987677713993208
MGT30	0.0963270909753563
SIO30	0.0959640847238615
PSV23	0.0881659101448704

Figura 2. Porcentaje de Cambio de la media de VB con respecto a la Pandemia.

Circuito o Barra	Porcentaje de Cambio Media VC
BUM30	6.787612113074843
BQE30	1.0
MAZ23	0.9657596405261542
VIR23	0.961629484399686
HER46	0.8008152411699722
IRR23	0.7018846364250224
HER40	0.6635044619824898
MNA30	0.5798060964659081
NRA30	0.5493989084675649
PUR46	0.4953747725977732
MNA23	0.3984527891864586
PSO10	0.3676152843812229
PSO23	0.3601761102291601
FEL23	0.2342696698177561
MAN23	0.2268527004515113
QHI30	0.2099684835738348
DOR10	0.1880332823290594
GRO23	0.1667677183183265
AMR23	0.1505305072396139
EDE30	0.1221808440556086
PSV23	0.0995195785425089
VIR30	0.0985117527144341
BOA23	0.0976647582444261
SIO30	0.095619312107746
MTT23	0.0885489536666666
AGU23	0.0834743038350893

Figura 3. Porcentaje de Cambio de la media de VC con respecto a la Pandemia.

Circuito o Barra	Porcentaje de Cambio Media IA
VIR23	0.9997397459633038
MAN23	0.9802470739493336
ARM10	0.9252359168745564
MAN10	0.8934959499746598
VBO10	0.8143951945277985
PSO10	0.8088351874211751
QHI30	0.7044616133978419
EDE30	0.6569466151884373
ELA23	0.6280520746592191
MNA30	0.568135974877564
BEO23	0.4936585462901253
MNA23	0.4758929872363838
VIR30	0.4745037292572059
PUR46	0.4537210199013653
NSA30	0.4458624709295665
VCT10	0.4449817051903073
EDE10	0.3729964312868392
VBO40	0.34897965277139
RIO40	0.3418170123675512
SLM30	0.3313835039978531
AZA10	0.2895119777107618
PUR40	0.2744643151601311
MTO10	0.2707488000032553
RIO30	0.2657717026494165

Figura 4. Porcentaje de Cambio de la media de IA con respecto a la Pandemia.

Circuito o Barra	Porcentaje de Cambio Media IB
PSO10	1.0506704540196825
VIR23	0.999261270813602
MAN23	0.9787669794178092
QHI30	0.8651837137258661
MAN10	0.8563968330080399
EDE30	0.6228172822245481
NSA30	0.6108173197865951
MNA30	0.5703673220395261
VIR30	0.5055680343693767
MNA23	0.4618699490350808
PUR46	0.457199419630782
DOR10	0.4281456819429708
DON23	0.3851354130662606
ESM23	0.349662575864528
SLM30	0.3368423045571277
RIO40	0.3301984733897544
VBO40	0.3240414267369483
PUR40	0.3204941569104832
VCT10	0.2865393912064558
SLM40	0.2607831838854671
RIO30	0.2604821458594787
ELA23	0.2458218726141593
BEO23	0.2408592538474589
VBO30	0.2240173634886969
MAR30	0.2165660553352102
IRR40	0.2043758707536951

Figura 5. Porcentaje de Cambio de la media de IB con respecto a la Pandemia.

Circuito o Barra	Porcentaje de Cambio Media IC
PSO10	1.4097656528970957
VIR23	0.9994030766775104
MAN23	0.9797649633535126
ARM10	0.9488060816700676
QHI30	0.9427495928892614
MAN10	0.8483747710599296
EDE30	0.6669538720364792
MNA30	0.5545374455082596
VIR30	0.5006459350881034
MNA23	0.4712691641525145
PUR46	0.4534228230150282
NSA30	0.4226576476727507
VBO40	0.4105856561746833
RIO40	0.3518565193044787
SLM30	0.3374387559504846
PUR40	0.332019926270827
ESM23	0.3281290024074009
RIO30	0.2725976100123093
SLM40	0.2725750510125244
VBO30	0.2538448845479863
DON23	0.2526573114396203
IRR40	0.2244118623639753
RIO10	0.2033552750035368
DOR10	0.2012703002709966
MAZ40	0.1912135729501304
INS23	0.1820627602128057

Figura 6. Porcentaje de Cambio de la media de IC con respecto a la Pandemia.

Circuito o Barra	Porcentaje de Cambio Media P
EDE30	124360.81346826578
CAM30	838.6677389669736
MAN40	221.55351031078865
MNA30	187.29905135005464
FLR23	167.75034975334117
PRO23	113.8470904625416
DDO23	98.81695251837829
LMC23	98.4492052810839
SIO23	91.66844033281625
AMR30	84.44297616782083
SUP23	73.0669638120332
VBO30	50.38669407224573
RSA23	46.95822314375921
AMR23	42.05004591132118
MTT23	38.43983253904544
PSV23	20.23418884759214
FIL23	17.780184006597384
ELA23	16.992127247594997
BUM30	16.426260233439912
FEL23	16.285693254120496
GTO23	13.81161339079511
REG40	13.665426630427694
CHI30	13.313625686074
COL23	13.123512873462706
SCH23	13.06355990020573
MTO23	11.70792067505871
AGU23	10.247567405839495

Figura 7. Porcentaje de Cambio de la media de P con respecto a la Pandemia.

Circuito o Barra	Porcentaje de Cambio Media Q
CAM30	inf
EDE30	32037.499930807116
MAN40	12799.245837084058
LMC23	474.19465404111
VBO30	385.8894166687504
PRO23	379.9392768027041
SUP23	191.1899559034229
SIO23	129.63838699721282
MGT23	100.49945080106224
DDO23	83.10455032992168
AMR23	71.45008094565705
MTT23	55.48066517964046
RSA23	49.32930204876544
FLR23	34.20834247251819
AMR30	30.47055654259295
MTO23	29.309329071866905
PSV23	26.84789953168606
ELA23	25.48416316861444
MNA30	21.93653262203398
SCH23	18.266754626760488
COL23	17.129954291406506
GTO23	14.46569454292058
FEL23	14.032180816118752
SNA23	13.455753207207488
AGU23	10.702186538305051
CAM23	8.734759668849257

Figura 8. Porcentaje de Cambio de la media de Q con respecto a la Pandemia.

Circuito o Barra	Porcentaje de Cambio Media VA
MTO30	2664455.7446765187
MAN40	61.03598301854248
VBO30	48.83723305876399
MAN30	27.813091347985356
PSO23	17.726785271962562
BEL23	10.657345251439144
IRR10	6.691139606341371
QHI30	5.22261491205905
IRR40	4.309875257089019
MNA30	3.9851520422740334
IRR30	3.3422223457668814
VMA30	3.1847302123803933
VMA23	2.93620181151382
VBO23	2.844579723269246
GRO23	2.7078276166308477
RIO40	2.5595979262233537
LMC30	2.325757062235123
AGU30	1.8858814309859169
ENE23	1.80161374962883
PSV30	1.565668498397437
PUR10	1.3856592338199445
SUP23	1.3438478056526495
DDO30	1.3361467524177166
DOR40	1.2990008204281565
BOA23	1.286625297428941
IES46	1.2545045093423377
DON23	1.1864410307404427
IES10	1.1441351038526335
NSA23	1.0404782066527372
EDE10	1.0301222945997142
MAZ23	0.9998090380276964
HER40	0.9986142876915138
LMC23	0.9945766880591812
PSO40	0.9939514281057106
BQE30	0.991068912670098
PSV23	0.9861683022606292
MAZ40	0.9776165272202084
SCH30	0.9739348439832526
PRO30	0.9687097690396866
ENE40	0.9558207406627248
MAN23	0.9256778202041188

Figura 9. Porcentaje de Cambio de la desviación estándar de VA con respecto a la Pandemia.

Circuito o Barra	Porcentaje de Cambio Media VB
MAN40	126.2833892942626
VBO30	53.013455136161845
MAN30	28.92693194507901
PSO23	17.810516714555643
IRR10	6.875696294549887
QHI30	5.288705809573215
BEL23	3.729988041273292
IRR30	3.4008473831371013
MNA30	3.3165389791900157
VMA30	3.2723219182275822
IRR40	3.099123850983005
VMA23	3.060966975890692
GRO23	2.81307797605406
VBO23	2.812978155959837
RIO40	2.594093982248852
AGU30	2.0076480846369025
LMC30	1.8978018071321463
DON23	1.7879978132541845
PSV30	1.5986243843474477
DDO30	1.3102703994191198
DOR40	1.3067941131904253
IES46	1.2494090056470215
BOA23	1.2116862480122064
AMR30	1.1992931265161568
IES10	1.0954327984162195
LMC23	1.0330288010978632
PRO30	1.0305909278358028
BQE30	1.0
MAZ23	0.9998079409128304
HER40	0.9986267446568086
PSO40	0.9934779986046772
MAZ40	0.9777707171682049
ROS40	0.9679449438860632
PSV23	0.9407374775956566
MAN23	0.9260637667398264
CHI30	0.9175954205006148
ROS30	0.9122167452739626
CHA23	0.90516733781821
PRA30	0.8811242796361528
PSO30	0.8713419312096884
NSA23	0.8592223648590595

Figura 10. Porcentaje de Cambio de la desviación estándar de VB con respecto a la Pandemia.

Circuito o Barra	Porcentaje de Cambio Media VC
VBO30	63.99515452758844
MAN30	29.47038592620357
BEL23	10.0990121579627
IRR10	7.24047932473632
QHI30	5.135205583276908
PSO23	5.051377201091124
MAN40	4.194370278696818
VMA30	3.323987657424145
IRR30	3.306599029018927
MNA30	3.3028937735639188
VMA23	3.177858805003547
VBO23	2.9423689431992712
GRO23	2.8884225719974648
IRR40	2.679853701629039
RIO40	2.573989376467134
AGU30	2.307981726939534
LMC30	1.9870692846369
DON23	1.6638620998089602
PSV30	1.5989611857182848
SUP23	1.3380756789921582
DDO30	1.3361313615390902
BOA23	1.2630908908342235
IES46	1.2550213592868205
IES10	1.1327062307282698
LMC23	1.0829298467610535
PSV23	1.0122367180040526
PRO30	1.0101343192535066
BQE30	1.0
MAZ23	0.9998088645173472
HER40	0.9986228762322484
PSO40	0.9938685769707776
MAZ40	0.9779161351096728
ENE40	0.9775180111781324
MAN23	0.925880614290592
CHI30	0.9238490766724692
ROS30	0.9132147577795424
CHA23	0.9082702990718796
NSA23	0.8673617301394008
RSA23	0.8169510012916855
PRA30	0.7898786805539153
ARM30	0.7567215831117028

Figura 11. Porcentaje de Cambio de la desviación estándar de VC con respecto a la Pandemia.

Circuito o Barra	Porcentaje de Cambio Media IA
IRR10	9.667484044777195
MNA30	1.3468001621154593
EDE30	1.3033281487557562
QHI30	1.0695081224068796
VIR23	0.9974059514662456
MAN23	0.9952759405680242
RSA23	0.7361687760160084
MAN10	0.5910234352753474
RIO40	0.577241057597966
EDE10	0.5459592727270948
FLR30	0.4620304291795709
ARM10	0.4276521520112768
NSA30	0.4232785083647788
RIO30	0.4129761380258518
PSO10	0.4070643340029918
VCT10	0.4060005138997121
SLM40	0.3693585409819969
MAR30	0.3651620866044374
FEL23	0.3195676074910168
MNA23	0.3194455160478162
ELA23	0.3142422518219222
MTO10	0.3030061780644371
VIR30	0.2503487675286254
AZA10	0.2386329548601779
VBO30	0.2339599914230294
BQE30	0.2319173276667252

Figura 12. Porcentaje de Cambio de la desviación estándar de IA con respecto a la Pandemia.

Circuito o Barra	Porcentaje de Cambio Media IB
IRR10	10.504092100293024
RSA30	1.5101749256858248
MNA30	1.3964263723496226
EDE30	1.268018551909726
QHI30	1.1448256291935253
MAN23	0.9950262374540016
VIR23	0.9948979614854452
PSO30	0.9934385238272044
FLR30	0.6292183339438898
RIO40	0.5917990659677259
NSA30	0.5253664058041972
MAN10	0.5131550839560246
REG10	0.4752628647288589
MAR30	0.4159769103268037
RIO30	0.4128478908639625
ELA23	0.387750225517926
SLM40	0.37599641153987
MNA23	0.3265218529876107
VIR30	0.2783927511865428
VCT10	0.2380589561718639
SIO30	0.2371280325428972
FLR23	0.2318269175537592
PSO10	0.2283171406520686
FEL23	0.2202289374122307
PRO30	0.2124644891159317
VBO30	0.2086698162323357

Figura 13. Porcentaje de Cambio de la desviación estándar de IB con respecto a la Pandemia.

Circuito o Barra	Porcentaje de Cambio SD IC
IRR10	9.67663111185398
RSA30	1.5272471942706942
MNA30	1.332230705333226
EDE30	1.2994776090906424
QHI30	1.1896098929616978
MAN23	0.9950751555077616
VIR23	0.9939473414533824
MAR30	0.6237675911927807
RIO40	0.5850379193645586
REG10	0.5489366974469392
MAN10	0.5189628728459531
RSA23	0.5067006789688768
DON23	0.477603043687111
RIO30	0.4205601516783711
FLR30	0.416952876641663
ARM10	0.4147888684449853
FEL23	0.3981267697240033
SLM40	0.370921945544729
NSA30	0.3417675589980601
RIO10	0.2875799064217212
FLR23	0.2667647422558045
VIR30	0.254851419427018
MNA23	0.2509349518016517
SIO30	0.247016560084432
VBO30	0.2458430834495383
GTO30	0.2408232922243103
MTO10	0.2162159465023095

Figura 14. Porcentaje de Cambio de la desviación estándar de IC con respecto a la Pandemia.

Circuito o Barra	Porcentaje de Cambio SD P
EDE30	400146.4922524784
VBO30	3200.769964380711
MNA30	1874.4773562010737
CAM30	1577.974988160978
MAN40	842.4972636679659
ELA23	409.0285874691321
SCH23	344.64165714902373
COL23	280.950886767416
GTO23	228.2469344165401
FEL23	203.13259551719003
AMR23	165.53936529816664
PSV23	130.90572657500945
AGU23	126.30685112045326
BOL23	120.33746198876644
CAM23	117.2175557541252
SNA23	112.96120960523156
BOA23	104.46471502656443
MTT23	98.7478381196036
SUP23	88.68899770575298
RSA23	86.59289985022812
FLR23	75.97516621106344
SIO23	75.11745069606954
LMC23	71.57048659351109
PRO23	69.88787166346923
DDO23	67.17552526672294
IRR23	63.11199287781422

Figura 15. Porcentaje de Cambio de la desviación estándar de P con respecto a la Pandemia.

Circuito o Barra	Porcentaje de Cambio SD Q
CAM30	inf
EDE30	271448.8929331579
VBO30	11367.730050604752
MAN40	4316.604334494265
ELA23	458.00949213068327
SCH23	319.5621488646702
MNA30	266.52238403797656
COL23	200.20521990733363
FEL23	183.1739040234034
GTO23	166.4641722046149
AMR23	163.12885954106304
SNA23	154.6525618538017
CAM23	131.3499638459167
AGU23	120.77859395676882
FLR23	108.1852736715816
MTT23	98.35940062991682
BOA23	96.80211247270324
SUP23	96.02628195460724
BOL23	93.36140528811563
PSV23	92.02000044121272
RSA23	79.75471885210324
LMC23	71.71992853403148
SIO23	69.93528754580403
DDO23	69.56918220755878
PRO23	67.81779630298263
AMR30	12.611901448004309

Figura 16. Porcentaje de Cambio de la desviación estándar de Q con respecto a la Pandemia.

De las anteriores tablas se puede ver claramente que el consumo de energía y los valores de voltaje cambiaron considerablemente con respecto a la pandemia; más sin embargo, el cambio de las corrientes fue mínimo, entonces inicialmente se pensaría descartar el histórico de voltajes y potencias durante la pandemia a la hora de entrenar un modelo de Machine Learning o Deep Learning, manteniendo los valores de corriente para este periodo de tiempo; más sin embargo, es importante notar que si se van a descartar los datos del periodo de pandemia o no, esto se debe hacer para todas las variables; ya que entre estas variables hay ciertas dependencias que el modelo de ML o DL debe aprender a codificar, es decir, la idea es que el modelo de ML o DL entrenado sea un modelo multivariable; por tanto se

hace necesario entrenar los modelos de ML o DL en dos posibles escenarios: Incluyendo los datos de la pandemia y sin incluir los datos de la pandemia, para así mirar en cuál de los dos escenarios se obtiene un mejor rendimiento.

Circuitos con problemas para un modelo de imputación:

Durante el análisis realizado para estudiar los casos posibles que deban ser identificados durante el modelo de reglas impuestas, se pudo identificar varios circuitos que podrían presentar problemas, a la hora de realizar algún modelo de imputación, debido a que no presentan un histórico fiable, debido a que prácticamente todos sus datos están perdidos, o son valores total o en su mayoría constante, lo cual deja un margen muy grande de saturación a la hora de aplicar un modelo, para estos casos se realizó un archivo Excel donde quedan condensados estos casos para que sean estudiados por el equipo de Chec, en estos claramente se pueden identificar algunos circuitos de prueba los cuales no tendrían problema puesto que como se acordó no serán tenidos en cuenta para aplicar los modelos de imputación, más sin embargo la cantidad de estos es mínima comparado con el total de circuitos.

SUBESTACIÓN	CIRCUITOS DESCARTADOS	CANTIDAD
IRR30	IRR30LH	1
IRR40	IRR40B29	1
LMC30	LMC30L13	1
AMR30	AMR30L13	1
MAN10	MAN10G02	1
MAN40	MAN40B20	1
ARM25	ARM25L12	1
MGT30	MGT30L12	1
ARM32	ARM323T1	1
MTO10	MTO10G02	1
NRA30	NRA30L11	1
NSA30	NSA30L13	1
ARM40	ARM40B20, ARM40S18, ARM40S19	3
PSO10	PSO10G02	1
QHI30	QHI30L12	1
REG10	REG10G02	1
RIO10	RIO10G02	1
AZA10	AZA10G02	1
RIO30	RIO30L12	1
ROS10	ROS10G02	1
ROS23	ROS23L18	1
ROS30	ROS30S16, S17, S25	3
SLM10	SLM10G02	1
SUP30	SUP30L13	1
SY'S_N	PRUEBA	1
TEB30	NO HAY DATOS	1
TEB40	NO HAY DATOS	1
TEST	PRUEBA	1
AZU23	AZU23L14	1
AZU30	AZU30L11	1
VBO70	VBO70L16	1
VCT10	VCT10G01, VCT10G02	2
2MZ2M	2MZ2MZ41	1
2MZ3M	2MZ3ML51	1
2NR2N	2NR2NR41	1
BQE10	BQE10G02, BQE10T01	2
2NS3N	2NS3NS41	1
BUM23	BUM23L15	1
2PR12	2PR12R01	1
CHA10	CHA10G02	2
2PR5S	2PR5SC21	1
2PS6P	2PS6PS81	1
CHI10	CHI10G02	2
2QHIQ	2QHIQH21	1
CHI23	CHI23L12, CHI23T20	9
2RI2R	2RI2RI31	2
2RI4S	2RI4SU51	1

Figura 17. Circuitos sin información

Cabe recalcar que la figura anterior solo tiene algunos circuitos para visualizar todo el informe, se enviara un archivo Excel al equipo de chec para que sean analizados y se tome una decisión con respecto a ellos.

Circuitos cruzados:

Tal cual se pudo evidenciar, en el análisis nodal de ejemplo realizado con el equipo chec, otro posible problema que se podría presentar al estudiar las bases de datos, es el caso donde hayan medidas cruzadas, es decir que la sumatoria de corrientes no coincida con su fase correspondiente, al realizar un estudio detallado no se pudieron evidenciar muchas situaciones así mas sin embargo si hubieron casos puntuales que se ven condensados en la siguientes figuras, y que por ende deben ser reportadas:

CRUZADOS				
MTT23				
BOL23				

Figura 18. Circuitos cruzados

CIRCUITO	TIEMPO_AJUSTADO	IA	IB	IC
MTT23L12	2020-01-01 00:00:00.000	8.0000000	4.0000000	8.0000000
MTT23L13	2020-01-01 00:00:00.000	17.0000000	20.0000000	31.0000000
MTT23L14	2020-01-01 00:00:00.000	1.0000000	5.0000000	2.0000000
MTT23T11	2020-01-01 00:00:00.000	40.0000000	26.0000000	23.0000000
MTT23L12	2020-01-01 00:15:00.000	8.0000000	4.0000000	8.0000000
MTT23L13	2020-01-01 00:15:00.000	17.0000000	20.0000000	31.0000000
MTT23L14	2020-01-01 00:15:00.000	1.0000000	5.0000000	2.0000000
MTT23T11	2020-01-01 00:15:00.000	40.0000000	26.0000000	23.0000000

Figura 19. Circuitos cruzados [MTT23]

CIRCUITO	TIEMPO_AJUSTADO	IA	IB	IC
BOL23T11	2021-11-25 10:15:00.000	21.0000000	16.0000000	16.0000000
BOL23L14	2021-11-25 10:15:00.000	3.0000000	4.0000000	1.0000000
BOL23L13	2021-11-25 10:15:00.000	6.0000000	7.0000000	9.0000000
BOL23L12	2021-11-25 10:15:00.000	8.0000000	13.0000000	7.0000000
BOL23T11	2021-11-25 10:00:00.000	21.0000000	16.0000000	16.0000000
BOL23L14	2021-11-25 10:00:00.000	3.0000000	4.0000000	1.0000000
BOL23L13	2021-11-25 10:00:00.000	6.0000000	7.0000000	9.0000000
BOL23L12	2021-11-25 10:00:00.000	8.0000000	13.0000000	7.0000000

Figura 20. Circuitos cruzados [BOL23]

En reuniones con el equipo Chec, al plantear el modelo de imputación, se propuso una estructura como sigue:

Pipeline Imputación de Datos

El siguiente es el pipeline propuesto para imputar las variables VA, VB, VC, IA, IB, IC, P y Q tanto del sistema SCADA ABB como del sistema SCADA SURVALENT:

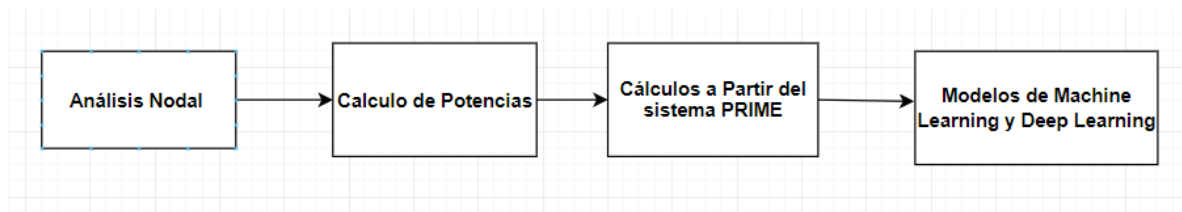


Figura 21. Pipeline de Imputación de Datos

Donde la idea es arrancar por un modelo de reglas impuestas, que incluye 3 bloques, en dicho orden de realización.

- **Análisis Nodal:** Consiste en completar información del SCADA, analizando los circuitos por subestación para los circuitos de cabecera, tomando valores de corriente de entrada y salida, más un análisis por nodo de los voltajes por fase, completando información por teoremas de Kirchhoff.
- **Cálculo de Potencias:** Consiste en completar información de las potencias Reactiva y activa, además de corregir la calidad de dichas magnitudes, teniendo en cuenta una comparación de potencias Aparentes, asumiendo un margen de error definido por el histórico de las potencias.
- **Cálculos a partir del sistema PRIME:** Este apartado no es implementado aún, debido a que no se suministrado la información completa del prime, pero en caso de que se encuentre disponible, la idea seria aprovechar las relaciones entre ambos sistemas para complementar información de las magnitudes eléctricas.

Teniendo en cuenta la estructura anterior, partimos a explicar cómo esta estipulado los dos primeros bloques de la arquitectura:

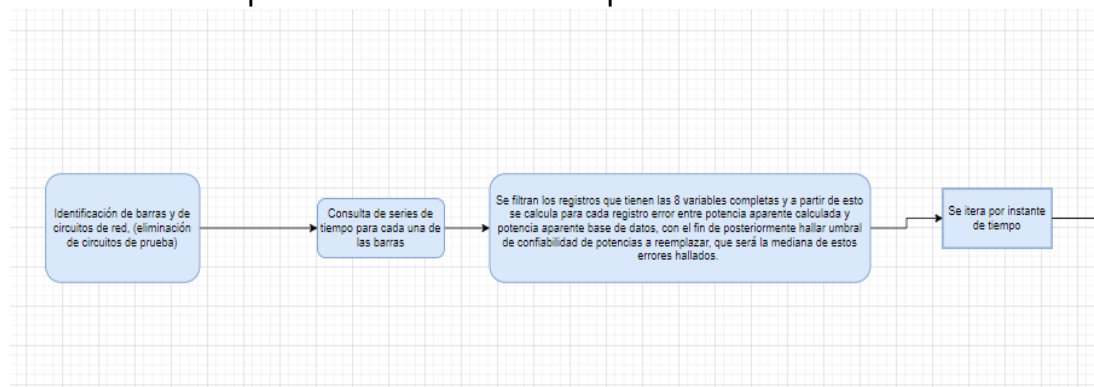
Descripción Algoritmo Análisis Nodal y Calculo de Potencias

(Modelo inicial de reglas impuestas)

“La explicación completa como diagrama de flujo, se encuentra definida en el siguiente enlace, la idea es ir modificando el esquema a medida que vamos agregando los bloques faltantes del modelo”.

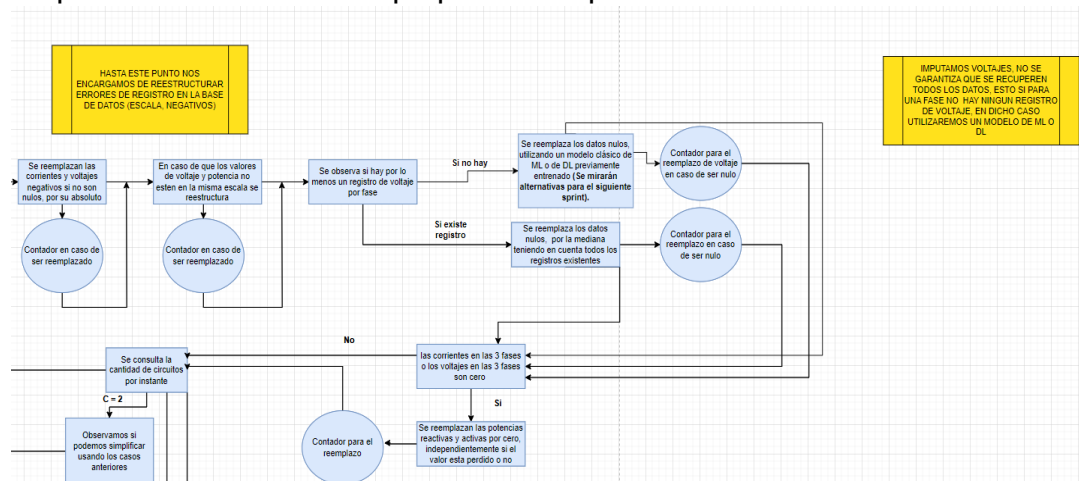
<https://drive.google.com/file/d/1-opOSHv3ZQ-Zs-PVwjpQb7h5lrMq5Z05/view?usp=sharing>

1. En primer lugar, se identifican las barras y circuitos de red presentes en el sistema; descartando aquellos circuitos cuyos códigos no coincidan con la nomenclatura estándar para un circuito de red y un circuito cabecera; ya que aquellos circuitos son circuitos de prueba.
2. Una vez obtenidos los códigos de los circuitos de red y de las barras presentes en el sistema, se procede a hacer de manera cíclica la consulta de las series de tiempo para cada circuito de red y para los circuitos pertenecientes a las respectivas barras.
3. Una vez obtenidas las series de tiempo para cada uno de los circuitos de la barra o para el respectivo circuito de red, se filtran aquellos registros en los cuales las 8 variables de interés (VA,VB,VC,IA,IB,IC,P,Q) están completas, de decir, no tienen datos nulos; esto con el fin de posteriormente en cada uno de estos registros calcular un porcentaje de error entre potencia aparente calculada a partir de corrientes y voltajes y potencia aparente calculada a partir de P y Q; para posteriormente, establecer la mediana de estos errores como umbral de confiabilidad de potencias a reemplazar.
4. Posteriormente, se itera por cada uno de los instantes de tiempo de las series de tiempo de los circuitos analizados; para hacer los respectivos análisis de acá en adelante por cada instante de tiempo.

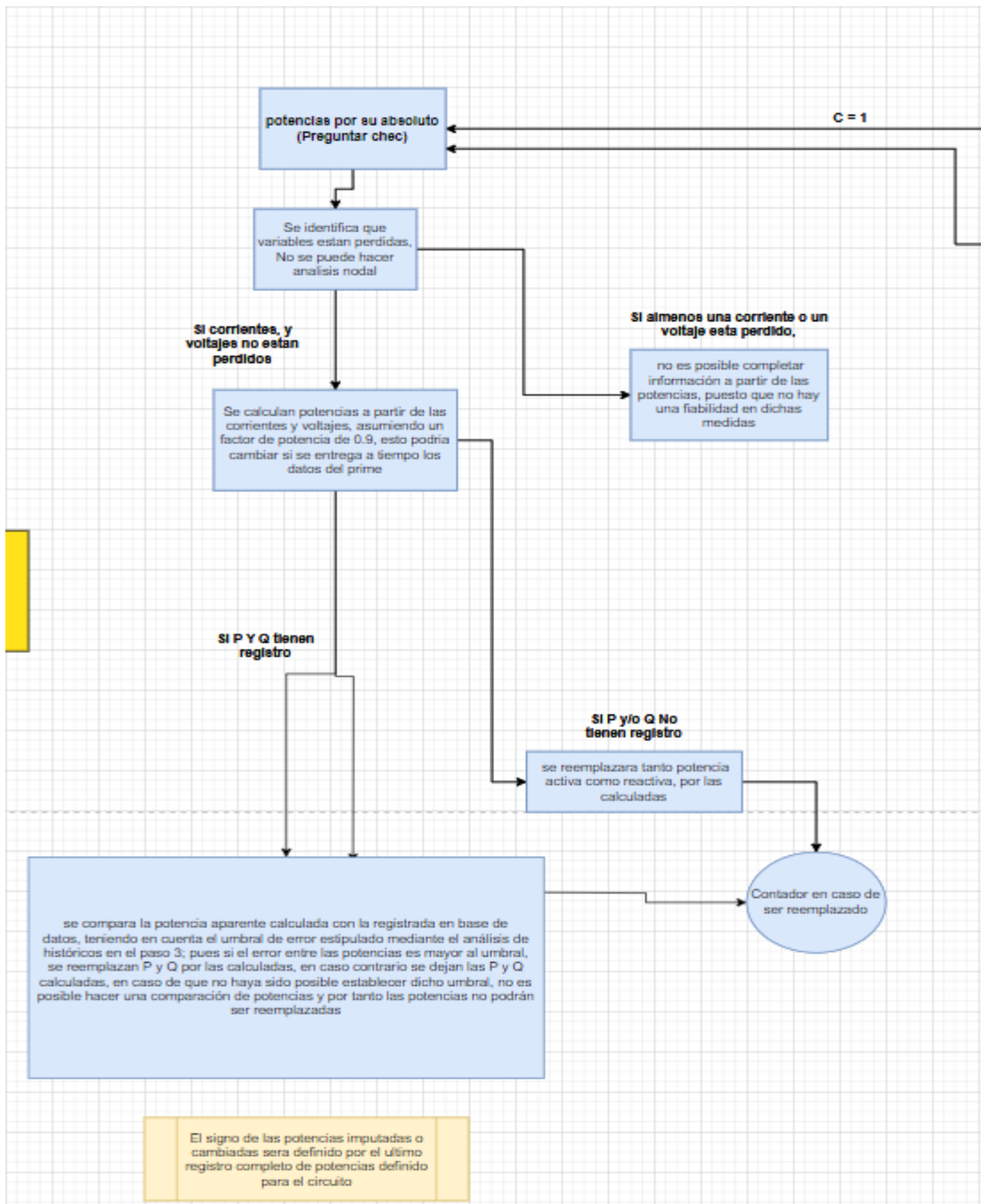


5. Se reemplazan los valores negativos de voltajes y corrientes por positivos; ya que estos valores siempre deben ser positivos (Contando en cuantos casos se hace esto para posteriormente sacar una estadística).

6. Debido a que se observó que en algunos registros los voltajes y potencias no estaban en escalas correctas, se procede a llevar los voltajes que no estén en escala correcta a su respectiva correcta escala (Contando en cuantos casos se hace esto para posteriormente sacar una estadística).
7. Luego, se observa si hay por lo menos un registro de voltaje no NULL por fase, en el caso de que, si halla por lo menos un registro de voltaje por fase, los valores NULL pertenecientes a la misma fase de voltaje, se imputan con la mediana de los voltajes no NULL de la fase, en caso contrario, se entrenará un modelo de Machine Learning o de Deep Learning para hacer las respectivas imputaciones; **es de destacar que este modelo se entrenará para el sprint 4.** (Contando en cuantos casos se hace esto para posteriormente sacar una estadística).
8. Posteriormente, se procede a filtrar aquellos circuitos que tienen sus tres corrientes en cero o sus tres voltajes en cero; esto con el fin de establecer en cero las potencias activas y reactivas de aquellos circuitos que cumplan con esta condición. (Contando en cuantos casos se hace esto para posteriormente sacar una estadística).
9. Luego, se procede a calcular la cantidad de circuitos presentes en el respectivo instante de tiempo para la respectiva barra o circuito de red.



Caso 1 (Cabeceras con un solo circuito y circuitos de red)



10.1.1 En caso de que haya un solo circuito, se procede a llevar las respectivas potencias del circuito a valor absoluto.

10.1.2. Se procede a identificar que variables están NULL en el respectivo circuito.

10.1.3. En caso de que haya por lo menos un voltaje o una corriente NULL, no se puede completar o reemplazar potencias en el circuito, en caso

contrario se calculan potencias (P, Q, S) a partir de corrientes y voltajes, asumiendo un factor de potencia de 0.9 (Lo cual puede cambiar en el caso en el que se entregue a tiempo los datos del primer y dicho factor de potencia se pueda calcular desde esta fuente).

10.1.4. En caso de que algún P o Q tenga registro NULL, se reemplaza tanto potencia activa como reactiva por las potencias calculadas, en caso contrario, se compara la potencia aparente calculada con la registrada en base de datos, teniendo en cuenta el umbral de error estipulado mediante el análisis de históricos en el paso 3; pues si el error entre las potencias es mayor al umbral, se reemplazan P y Q por las calculadas, en caso contrario se dejan las P y Q calculadas, en caso de que no haya sido posible establecer dicho umbral, no es posible hacer una comparación de potencias y por tanto las potencias no podrán ser reemplazadas. Es de destacar que el signo de las potencias imputadas o cambiadas será definido por el signo del último registro completo de potencias en base de datos para el respectivo circuito; ya que estos signos se han ido corrigiendo en el sistema, mientras que, si no existe un último registro completo en base de datos, dichas potencias quedarán con valor positivo (Contando en cuantos casos se hace esto para posteriormente sacar una estadística).

Caso (2 Circuitos):

10.2.1. En el caso en el que hallan dos circuitos, se procede a mirar si alguno de esos circuitos tiene todas sus corrientes en cero; pues este escenario quiere decir que este circuito no tiene influencia en el análisis nodal; por tanto, se procede a hacer todo el análisis del caso en el cual solo hay un circuito, pero solo con el circuito que no tiene todas sus corrientes en cero. En caso contrario, se procede a analizar cuantas parejas de corrientes en sus respectivas fases es posible comparar, es decir, aquellas fases de corriente en las cuales no hay valores NULL.

10.2.2. En caso de que no haya ninguna pareja de corrientes que se pueda comparar, no se podrá hacer análisis nodal; ya que no es posible saber si ambos circuitos son de entrada o ambos circuitos son de salida (posiblemente teniendo un circuito que hace parte de la sección de la barra para el cual no hay medida registrada) o hay un circuito de entrada y otro de salida.

En caso de que haya una pareja de corrientes comparable, se calcula el porcentaje de diferencia entre dichas corrientes y en caso de que ese porcentaje de diferencia sea mayor al umbral definido, se genera una alerta; ya que eso quiere decir que ambos circuitos están entrando o ambos circuitos están saliendo y por tanto hay un tercer circuito para el cual no está llegando registro (Contando en cuantos casos se hace esto para posteriormente sacar una estadística), mientras que si ese porcentaje de diferencia entre dichas corrientes

es menor o igual al umbral definido; eso quiere decir que un circuito está entrando y el otro está saliendo, en ese caso, se imputan las corrientes NULL que se puedan imputar (Contando en cuantos casos se hace esto para posteriormente sacar una estadística).

En caso de que haya dos parejas de corrientes comparables, se calcula el porcentaje de diferencia entre las corrientes para cada una de las parejas, en caso de que ese porcentaje de diferencia sea mayor al umbral definido para alguna de las parejas, se genera una alerta; ya que eso quiere decir que ambos circuitos están entrando o ambos circuitos están saliendo y por tanto hay un tercer circuito para el cual no está llegando registro (Contando en cuantos casos se hace esto para posteriormente sacar una estadística), mientras que si ese porcentaje de diferencia entre dichas corrientes para ambas parejas es menor o igual al umbral definido; eso quiere decir que un circuito está entrando y el otro está saliendo, en ese caso, se imputan las corrientes NULL que se puedan imputar (Contando en cuantos casos se hace esto para posteriormente sacar una estadística).

En caso de que haya tres parejas de corrientes comparables, se calcula el porcentaje de diferencia entre las corrientes para cada una de las parejas, en caso de que ese porcentaje de diferencia sea mayor al umbral definido para alguna de las parejas, se genera una alerta; ya que eso quiere decir que ambos circuitos están entrando o ambos circuitos están saliendo y por tanto hay un tercer circuito para el cual no está llegando registro (Contando en cuantos casos se hace esto para posteriormente sacar una estadística).

10.2.3. En caso de que haya por lo menos un voltaje o una corriente NULL, no se puede completar o reemplazar potencias en el circuito, en caso contrario se calculan potencias (P, Q, S) a partir de corrientes y voltajes, asumiendo un factor de potencia de 0.9 (Lo cual puede cambiar en el caso en el que se entregue a tiempo los datos del primer y dicho factor de potencia se pueda calcular desde esta fuente).

10.2.4. En caso de que algún P o Q tenga registro NULL, se reemplaza tanto potencia activa como reactiva por las potencias calculadas, en caso contrario, se compara la potencia aparente calculada con la registrada en base de datos, teniendo en cuenta el umbral de error estipulado mediante el análisis de históricos en el paso 3; pues si el error entre las potencias es mayor al umbral, se reemplazan P y Q por las calculadas, en caso contrario se dejan las P y Q calculadas, en caso de que no haya sido posible establecer dicho umbral, no es posible hacer una comparación de potencias y por tanto las potencias no podrán ser reemplazadas. Es de destacar que el signo de las potencias imputadas o cambiadas será definido por el signo del último registro completo de potencias en base de datos para el respectivo circuito; ya que estos signos se han ido corrigiendo en el sistema, mientras que si no

existe un último registro completo en base de datos, dichas potencias quedarán con valor positivo (Contando en cuantos casos se hace esto para posteriormente sacar una estadística).

Caso (3 Circuitos):

10.3.1. En el caso en el que hallan tres circuitos, se procede a mirar si alguno de esos circuitos tiene todas sus corrientes en cero; pues este escenario quiere decir que este circuito no tiene influencia en el análisis nodal; por tanto, se procede a hacer todo el análisis del caso en el cual solo hay dos circuitos, pero solo con los dos circuitos que no tienen todas sus corrientes en cero. En caso contrario, se procede a analizar cuantas ternas de corrientes en sus respectivas fases es posible comparar, es decir, aquellas fases de corriente en las cuales no hay valores NULL.

10.3.2. En caso de que no haya ninguna terna de corrientes que se pueda comparar, no se podrá hacer análisis nodal; ya que no es posible saber si los tres circuitos son de entrada o los tres circuitos son de salida (posiblemente teniendo un circuito que hace parte de la sección de la barra para el cual no hay medida registrada) o hay circuitos de entrada y de salida.

En caso de que haya una terna de corrientes comparable (dos entrando y una saliendo (la mayor) o viceversa), se calcula el porcentaje de diferencia entre dichas corrientes y en caso de que ese porcentaje de diferencia sea mayor al umbral definido, se genera una alerta; ya que eso quiere decir que los tres circuitos están entrando o los tres circuitos están saliendo y por tanto hay un cuarto circuito para el cual no está llegando registro (Contando en cuantos casos se hace esto para posteriormente sacar una estadística), mientras que si ese porcentaje de diferencia entre dichas corrientes es menor o igual al umbral definido; eso quiere decir que hay circuitos entrando y circuitos saliendo, en ese caso, se imputan las corrientes NULL que se puedan imputar (Contando en cuantos casos se hace esto para posteriormente sacar una estadística).

En caso de que haya dos ternas de corrientes comparables (dos entrando y una saliendo (la mayor) o viceversa), se calcula el porcentaje de diferencia entre las corrientes para cada una de las ternas, en caso de que ese porcentaje de diferencia sea mayor al umbral definido para alguna de las parejas, se genera una alerta; ya que eso quiere decir que los tres circuitos están entrando o los tres circuitos están saliendo y por tanto hay un cuarto circuito para el cual no está llegando registro (Contando en cuantos casos se hace esto para posteriormente sacar una estadística), mientras que si ese porcentaje de diferencia entre dichas corrientes para las tres parejas es menor o igual al umbral definido; eso quiere decir que hay circuitos entrando y circuitos saliendo, en ese caso, se imputan las corrientes NULL que se puedan imputar (Contando en cuantos casos se hace esto para posteriormente sacar una estadística).

En caso de que haya tres parejas de corrientes comparables (dos entrando y una saliendo (la mayor) o viceversa), se calcula el porcentaje de diferencia entre las corrientes para cada una de las ternas, en caso de que ese porcentaje de diferencia sea mayor al umbral definido para alguna de las ternas, se genera una alerta; ya que eso quiere decir que los tres circuitos están entrando o los tres circuitos están saliendo y por tanto hay un cuarto circuito para el cual no está llegando registro (Contando en cuantos casos se hace esto para posteriormente sacar una estadística).

10.3.3 En caso de que haya por lo menos un voltaje o una corriente NULL, no se puede completar o reemplazar potencias en el circuito, en caso contrario se calculan potencias (P, Q,S) a partir de corrientes y voltajes, asumiendo un factor de potencia de 0.9 (Lo cual puede cambiar en el caso en el que se entregue a tiempo los datos del primer y dicho factor de potencia se pueda calcular desde esta fuente).

10.3.4 En caso de que algún P o Q tenga registro NULL, se reemplaza tanto potencia activa como reactiva por las potencias calculadas, en caso contrario, se compara la potencia aparente calculada con la registrada en base de datos, teniendo en cuenta el umbral de error estipulado mediante el análisis de históricos en el paso 3; pues si el error entre las potencias es mayor al umbral, se reemplazan P y Q por las calculadas, en caso contrario se dejan las P y Q calculadas, en caso de que no haya sido posible establecer dicho umbral, no es posible hacer una comparación de potencias y por tanto las potencias no podrán ser reemplazadas. Es de destacar que el signo de las potencias imputadas o cambiadas será definido por el signo del último registro completo de potencias en base de datos para el respectivo circuito; ya que estos signos se han ido corrigiendo en el sistema, mientras que si no existe un último registro completo en base de datos, dichas potencias quedarán con valor positivo (Contando en cuantos casos se hace esto para posteriormente sacar una estadística).

Caso n Circuitos

10.4.1. En el caso en el que hallan más de tres circuitos, se procede a analizar cuantas fases tienen sus corrientes completas, es decir, sin valores NULL.

10.4.2. En caso de que no haya ninguna fase con todas sus corrientes completas, no se podrá hacer análisis nodal; ya que no es posible saber si los circuitos son de entrada o son de salida (posiblemente teniendo un circuito que hace parte de la sección de la barra para el cual no hay medida registrada) o hay circuitos de entrada y de salida.

En caso de que haya una fase con todas sus corrientes completas, se analizan cuáles son todas las posibles combinaciones de circuitos entrando y saliendo, y para cada una de estas combinaciones se calcula el porcentaje de diferencia entre dichas corrientes y en caso de que ese porcentaje de diferencia sea mayor al umbral definido para todas las combinaciones, se genera una alerta; ya que eso quiere decir que todos los circuitos están entrando o están saliendo y por tanto hay un circuito para el cual no está llegando registro (Contando en cuantos casos se hace esto para posteriormente sacar una estadística), mientras que si ese porcentaje de diferencia entre dichas corrientes es menor o igual al umbral definido por lo menos para una de las combinaciones; eso quiere decir que hay circuitos entrando y circuitos saliendo, en ese caso, se imputan las corrientes NULL que se puedan imputar, tomando como referencia para el análisis nodal la combinación que dio el menor porcentaje de diferencia (Contando en cuantos casos se hace esto para posteriormente sacar una estadística).

En caso de que hallan dos fases con todas sus corrientes completas, se analizan cuáles son todas las posibles combinaciones de circuitos entrando y saliendo para ambas fases, y para cada una de estas combinaciones se calcula el porcentaje de diferencia entre dichas corrientes para ambas fases y en caso de que ese porcentaje de diferencia sea mayor al umbral definido para todas las combinaciones en por lo menos una fase, se genera una alerta; ya que eso quiere decir que todos los circuitos están entrando o están saliendo y por tanto hay un circuito para el cual no está llegando registro (Contando en cuantos casos se hace esto para posteriormente sacar una estadística), mientras que si ese porcentaje de diferencia entre dichas corrientes es menor o igual al umbral definido por lo menos para una de las combinaciones en cada una de las fases; eso quiere decir que hay circuitos entrando y circuitos saliendo, en ese caso, se imputan las corrientes NULL que se puedan imputar, tomando como referencia para el análisis nodal la combinación que dio el menor porcentaje de diferencia en cada una de las dos fases (Contando en cuantos casos se hace esto para posteriormente sacar una estadística).

En caso de que hallan tres fases con todas sus corrientes completas, se analizan cuáles son todas las posibles combinaciones de circuitos entrando y saliendo para las tres fases, y para cada una de estas combinaciones se calcula el porcentaje de diferencia entre dichas corrientes para las tres fases y en caso de que ese porcentaje de diferencia sea mayor al umbral definido para todas las combinaciones en por lo menos una fase, se genera una alerta; ya que eso quiere decir que todos los circuitos están entrando o están saliendo y por tanto hay un circuito para el cual no está llegando registro (Contando en cuantos casos se hace esto para posteriormente sacar una estadística).

10.4.3. En caso de que haya por lo menos un voltaje o una corriente NULL, no se puede completar o reemplazar potencias en el circuito, en caso contrario se calculan potencias (P, Q, S) a partir de corrientes y voltajes, asumiendo un factor

de potencia de 0.9 (Lo cual puede cambiar en el caso en el que se entregue a tiempo los datos del primer y dicho factor de potencia se pueda calcular desde esta fuente).

10.4.4 En caso de que algún P o Q tenga registro NULL, se reemplaza tanto potencia activa como reactiva por las potencias calculadas, en caso contrario, se compara la potencia aparente calculada con la registrada en base de datos, teniendo en cuenta el umbral de error estipulado mediante el análisis de históricos en el paso 3; pues si el error entre las potencias es mayor al umbral, se reemplazan P y Q por las calculadas, en caso contrario se dejan las P y Q calculadas, en caso de que no haya sido posible establecer dicho umbral, no es posible hacer una comparación de potencias y por tanto las potencias no podrán ser reemplazadas. Es de destacar que el signo de las potencias imputadas o cambiadas será definido por el signo del último registro completo de potencias en base de datos para el respectivo circuito; ya que estos signos se han ido corrigiendo en el sistema, mientras que, si no existe un último registro completo en base de datos, dichas potencias quedarán con valor positivo (Contando en cuantos casos se hace esto para posteriormente sacar una estadística).

El modelo de reglas impuestas, entrega un informe preliminar donde se especifica, de manera inicial cuantos datos habían perdidos antes de ingresar al modelo, y posterior a la ejecución de este por cada subestación se especifica cuantos datos se imputaron, bajo que método, y también casos puntuales donde no haya sido posible determinar un análisis nodal bajo supuestos de circuitos faltantes, la estructura del informe viene estipulada de la siguiente manera:

	VA_NULOS_INICIALES	VA_NULOS_FINALES	VB_NULOS_INICIALES	VB_NULOS_FINALES	VC_NULOS_INICIALES
TEB40	0	0	0	0	0
GTO23	0	0	0	0	0
HER40	0	0	0	0	0
ENE46	0	0	0	0	0
SCH30	0	0	0	0	0
---	---	---	---	---	---
PRO30	0	0	0	0	0
SLM40	0	0	0	0	0
ROS23	0	0	0	0	0
RIO40	0	0	0	0	0
SCH23	0	0	0	0	0

Figura 22. Estructura del informe

Resultados del algoritmo sobre circuitos de testeo

MTO30: Tomando 100 muestras

	CIRCUITO	TIEMPO AJUSTADO	IA	IB	IC	VA	VB	VC	P	Q
0	MTO30L10	2020-01-01 00:00:00	0E-7	0E-7	0E-7	None	None	None	0E-7	-0.0330000
1	MTO30L11	2020-01-01 00:00:00	41.0000000	45.0000000	46.0000000	34.3330000	33.8390000	34.2100000	-2.9100000	-0.2400000
2	MTO30L12	2020-01-01 00:00:00	21.0000000	21.0000000	23.0000000	None	None	None	1.0000000	-1.1100000
3	MTO30L13	2020-01-01 00:00:00	62.0000000	61.0000000	64.0000000	None	None	None	-3.4600000	0.1800000
4	MTO30L14	2020-01-01 00:00:00	61.0000000	57.0000000	56.0000000	None	None	None	-3.6200000	-0.0800000
5	MTO30T15	2020-01-01 00:00:00	144.0000000	144.0000000	152.0000000	None	None	None	None	None
6	MTO30L10	2020-01-01 00:15:00	0E-7	0E-7	0E-7	None	None	None	0E-7	-0.0350000
7	MTO30L11	2020-01-01 00:15:00	42.0000000	45.0000000	46.0000000	34.3330000	33.8390000	34.2100000	-2.8000000	-0.3500000
8	MTO30L12	2020-01-01 00:15:00	20.0000000	20.0000000	21.0000000	None	None	None	1.0000000	-1.0000000
9	MTO30L13	2020-01-01 00:15:00	61.0000000	57.0000000	61.0000000	None	None	None	-3.6800000	0.1000000
10	MTO30L14	2020-01-01 00:15:00	58.0000000	54.0000000	54.0000000	None	None	None	-3.5600000	-0.0800000
11	MTO30T15	2020-01-01 00:15:00	152.0000000	144.0000000	144.0000000	None	None	None	None	None
12	MTO30L10	2020-01-01 00:30:00	0E-7	0E-7	0E-7	None	None	None	0E-7	-0.0320000
13	MTO30L11	2020-01-01 00:30:00	43.0000000	47.0000000	48.0000000	34.3330000	33.9630000	34.2100000	-2.9600000	-0.5100000

Figura 23. Estructura inicial

MTO30	
VA_NULOS_INICIALES	83
VA_NULOS_FINALES	0
VB_NULOS_INICIALES	83
VB_NULOS_FINALES	0
VC_NULOS_INICIALES	83
VC_NULOS_FINALES	0
IA_NULOS_INICIALES	0
IA_NULOS_FINALES	0
IB_NULOS_INICIALES	0
IB_NULOS_FINALES	0
IC_NULOS_INICIALES	0
IC_NULOS_FINALES	0
P_NULOS_INICIALES	16
P_NULOS_FINALES	0
Q_NULOS_INICIALES	16

Figura 24. Tabla de resumen

Inicialmente teníamos para 100 muestras 83 datos perdidos para los voltajes, y 16 datos perdidos de potencias, finalmente mediante reglas impuestas podemos evidenciar que se recuperaron todos. Dentro del informe también es posible determinar mediante que método como se ve a continuación:

VA_IMPUTADO_MEDIANA	83
VA_IMPUTADO_NIVEL_TENSION	0
VB_IMPUTADO_MEDIANA	83
VB_IMPUTADO_NIVEL_TENSION	0
VC_IMPUTADO_MEDIANA	83
VC_IMPUTADO_NIVEL_TENSION	0
P_CAMBIADO_CERO	17
Q_CAMBIADO_CERO	17
P_CAMBIADO_POR_CALCULADO	49
Q_CAMBIADO_POR_CALCULADO	49
CABEZERAS_CON_CIRCUITO_SIN_REGISTRO	17

Figura 25. Métodos de imputación

	CIRCUITO	TIEMPO_AJUSTADO	IA	IB	IC	VA	VB	VC	\
0	MT030L10	2020-01-01 00:00:00	0.0	0.0	0.0	34.333	33.839	34.21	
1	MT030L11	2020-01-01 00:00:00	41.0	45.0	46.0	34.333	33.839	34.21	
2	MT030L12	2020-01-01 00:00:00	21.0	21.0	23.0	34.333	33.839	34.21	
3	MT030L13	2020-01-01 00:00:00	62.0	61.0	64.0	34.333	33.839	34.21	
4	MT030L14	2020-01-01 00:00:00	61.0	57.0	56.0	34.333	33.839	34.21	
..	
95	MT030T15	2020-01-01 03:45:00	120.0	120.0	120.0	34.333	33.963	34.21	
96	MT030L10	2020-01-01 04:00:00	0.0	0.0	0.0	34.333	33.963	34.21	
97	MT030L11	2020-01-01 04:00:00	39.0	43.0	44.0	34.333	33.963	34.21	
98	MT030L12	2020-01-01 04:00:00	22.0	22.0	22.0	34.333	33.963	34.21	
99	MT030L13	2020-01-01 04:00:00	51.0	48.0	51.0	34.333	33.963	34.21	

Figura 26. Estructura imputada

TEB40: Es un caso donde tenemos todos los valores perdidos, esto implica que no es posible hacer análisis nodal y tampoco es posible calcular potencias, mas sin embargo no se debe cortar la ejecución del algoritmo; pues este debe estar blindado para dichos casos:

	CIRCUITO	TIEMPO_AJUSTADO	IA	IB	IC	VA	VB	VC	P	Q
0	TEB40T21	2020-05-07 14:00:00	None	None	None	None	None	None	None	None
1	TEB40T21	2020-05-07 21:00:00	None	None	None	None	None	None	None	None
2	TEB40T21	2020-05-07 21:15:00	None	None	None	None	None	None	None	None
3	TEB40T21	2020-05-07 21:30:00	None	None	None	None	None	None	None	None
4	TEB40T21	2020-05-07 21:45:00	None	None	None	None	None	None	None	None
5	TEB40T21	2020-05-07 22:00:00	None	None	None	None	None	None	None	None
6	TEB40T21	2020-05-07 22:15:00	None	None	None	None	None	None	None	None
7	TEB40T21	2020-05-07 22:30:00	None	None	None	None	None	None	None	None

Figura 27. Datos iniciales

Si observamos el reporte dentro del algoritmo obtenemos lo siguiente:

TEB40	
VA_NULOS_INICIALES	100
VA_NULOS_FINALES	100
VB_NULOS_INICIALES	100
VB_NULOS_FINALES	100
VC_NULOS_INICIALES	100
VC_NULOS_FINALES	100
IA_NULOS_INICIALES	100
IA_NULOS_FINALES	100
IB_NULOS_INICIALES	100
IB_NULOS_FINALES	100
IC_NULOS_INICIALES	100
IC_NULOS_FINALES	100
P_NULOS_INICIALES	100
P_NULOS_FINALES	100
Q_NULOS_INICIALES	100
Q_NULOS_FINALES	100

Figura 28. Reporte

MAR30:

	CIRCUITO	TIEMPO AJUSTADO	IA	IB	IC	VA	VB	VC	P	Q
0	MAR30L12	2020-01-01 00:00:00	23.00000000	21.00000000	41.00000000	30.5379600	26.0641800	32.3838700	0.0040000	-0.6470000
1	MAR30L13	2020-01-01 00:00:00	42.00000000	44.00000000	41.00000000	32.5153500	32.5551400	32.5049700	1.4720000	1.5040000
2	MAR30L12	2020-01-01 00:15:00	23.00000000	21.00000000	41.00000000	30.5379600	26.0641800	32.3838700	0.0040000	-0.6470000
3	MAR30L13	2020-01-01 00:15:00	42.00000000	44.00000000	41.00000000	32.5153500	32.5551400	32.5049700	1.4720000	1.5040000
4	MAR30L12	2020-01-01 00:30:00	23.00000000	21.00000000	41.00000000	30.5379600	26.0641800	32.3838700	0.0040000	-0.6470000
5	MAR30L13	2020-01-01 00:30:00	42.00000000	44.00000000	41.00000000	32.5153500	32.5551400	32.5049700	1.4720000	1.5040000
6	MAR30L12	2020-01-01 00:45:00	23.00000000	21.00000000	41.00000000	30.5379600	26.0641800	32.3838700	0.0040000	-0.6470000
7	MAR30L13	2020-01-01 00:45:00	42.00000000	44.00000000	41.00000000	32.5153500	32.5551400	32.5049700	1.4720000	1.5040000
8	MAR30L12	2020-01-01 01:00:00	23.00000000	21.00000000	41.00000000	30.5379600	26.0641800	32.3838700	0.0040000	-0.6470000
9	MAR30L13	2020-01-01 01:00:00	42.00000000	44.00000000	41.00000000	32.5153500	32.5551400	32.5049700	1.4720000	1.5040000
10	MAR30L12	2020-01-01 01:15:00	23.00000000	21.00000000	41.00000000	30.5379600	26.0641800	32.3838700	0.0040000	-0.6470000

Figura 29. Datos iniciales

MAR30	
VA_NULOS_INICIALES	0
VA_NULOS_FINALES	0
VB_NULOS_INICIALES	0
VB_NULOS_FINALES	0
VC_NULOS_INICIALES	0
VC_NULOS_FINALES	0
IA_NULOS_INICIALES	0
IA_NULOS_FINALES	0
IB_NULOS_INICIALES	0
IB_NULOS_FINALES	0
IC_NULOS_INICIALES	0
IC_NULOS_FINALES	0
P_NULOS_INICIALES	0
P_NULOS_FINALES	0
Q_NULOS_INICIALES	0
Q_NULOS_FINALES	0

Figura 30. Reporte

P_CAMBIADO_POR_CALCULADO	50
Q_CAMBIADO_POR_CALCULADO	50
CABEZERAS_CON_CIRCUITO_SIN_REGISTRO	50
IA_IMPUTADO_POR_ANALISIS_NODAL	0
IB_IMPUTADO_POR_ANALISIS_NODAL	0
IC_IMPUTADO_POR_ANALISIS_NODAL	0

Figura 31. Registros cambiados

MGT30:

	CIRCUITO	TIEMPO_AJUSTADO	IA	IB	IC	VA	VB	VC	P	Q
0	MGT30L12	2020-01-01 00:00:00	33.00000000	35.00000000	36.00000000	0E-7	0E-7	0E-7	0E-7	0E-7
1	MGT30L13	2020-01-01 00:00:00	0E-7	0E-7	0E-7	32.5222700	32.7800400	32.0569000	0E-7	0E-7
2	MGT30T11	2020-01-01 00:00:00	30.00000000	36.00000000	36.00000000	32.5274600	32.9980200	32.4375000	1.3180000	0.5880000
3	MGT30L12	2020-01-01 00:15:00	33.00000000	35.00000000	36.00000000	0E-7	0E-7	0E-7	0E-7	0E-7
4	MGT30L13	2020-01-01 00:15:00	0E-7	0E-7	0E-7	32.5222700	32.7800400	32.0569000	0E-7	0E-7
5	MGT30T11	2020-01-01 00:15:00	30.00000000	36.00000000	36.00000000	32.5274600	32.9980200	32.4375000	1.3180000	0.5880000
6	MGT30L12	2020-01-01 00:30:00	33.00000000	35.00000000	36.00000000	0E-7	0E-7	0E-7	0E-7	0E-7
7	MGT30L13	2020-01-01 00:30:00	0E-7	0E-7	0E-7	32.5222700	32.7800400	32.0569000	0E-7	0E-7
8	MGT30T11	2020-01-01 00:30:00	30.00000000	36.00000000	36.00000000	32.5274600	32.9980200	32.4375000	1.3180000	0.5880000
9	MGT30L12	2020-01-01 00:45:00	33.00000000	35.00000000	36.00000000	0E-7	0E-7	0E-7	0E-7	0E-7
10	MGT30L13	2020-01-01 00:45:00	0E-7	0E-7	0E-7	32.5222700	32.7800400	32.0569000	0E-7	0E-7
11	MGT30T11	2020-01-01 00:45:00	30.00000000	36.00000000	36.00000000	32.5274600	32.9980200	32.4375000	1.3180000	0.5880000

Figura 32. Datos iniciales [Caso ideal]

MGT30	
VA_NULOS_INICIALES	0
VA_NULOS_FINALES	0
VB_NULOS_INICIALES	0
VB_NULOS_FINALES	0
VC_NULOS_INICIALES	0
VC_NULOS_FINALES	0
IA_NULOS_INICIALES	0
IA_NULOS_FINALES	0
IB_NULOS_INICIALES	0
IB_NULOS_FINALES	0
IC_NULOS_INICIALES	0
IC_NULOS_FINALES	0
P_NULOS_INICIALES	0
P_NULOS_FINALES	0
Q_NULOS_INICIALES	0
Q_NULOS_FINALES	0

Figura 33. Reporte

CORRECCION_ESCALA_Q	0
VA_IMPUTADO_MEDIANA	0
VA_IMPUTADO_NIVEL_TENSION	0
VB_IMPUTADO_MEDIANA	0
VB_IMPUTADO_NIVEL_TENSION	0
VC_IMPUTADO_MEDIANA	0
VC_IMPUTADO_NIVEL_TENSION	0
P_CAMBIADO_CERO	67
Q_CAMBIADO_CERO	67
P_CAMBIADO_POR_CALCULADO	0
Q_CAMBIADO_POR_CALCULADO	0
CABEZERAS_CON_CIRCUITO_SIN_REGISTRO	0
IA_IMPUTADO_POR_ANALISIS_NODAL	0
IB_IMPUTADO_POR_ANALISIS_NODAL	0
IC_IMPUTADO_POR_ANALISIS_NODAL	0

Figura 34. Reporte de cambios

ARM30:

	CIRCUITO	TIEMPO_AJUSTADO	IA	IB	IC	VA	VB	VC	P	Q
0	ARM30L13	2020-01-01 00:00:00	155.4832000	148.7688000	153.1785000	32.6085800	33.2280800	33.2929400	4.7661240	-7.5851730
1	ARM30L15	2020-01-01 00:00:00	116.2590000	113.9771000	131.0305000	32.7175200	33.6847800	33.1930500	3.9963290	-5.4963910
2	ARM30L18	2020-01-01 00:00:00	None	None	None	None	None	None	None	0.6550700
3	ARM30T11	2020-01-01 00:00:00	43.1684600	43.6573700	42.1932000	33.2073100	33.5337700	33.6650900	-2.1559140	1.4057380
4	ARM30T16	2020-01-01 00:00:00	312.3733000	302.8504000	302.9674000	32.8098800	32.9257200	33.3067600	17.6105900	1.9616500
5	ARM30T19	2020-01-01 00:00:00	None	None	None	None	None	None	None	0.6767570
6	ARM30L13	2020-01-01 00:15:00	150.3348000	148.9479000	148.7197000	32.6085800	33.2280800	33.2929400	4.7661240	-7.5851730
7	ARM30L15	2020-01-01 00:15:00	116.2590000	113.9771000	131.0305000	32.7175200	33.6847800	33.1930500	3.9963290	-5.2819770
8	ARM30L18	2020-01-01 00:15:00	None	None	None	None	None	None	None	0.6550700
9	ARM30T11	2020-01-01 00:15:00	43.1684600	43.6573700	42.1932000	33.2281100	33.5702200	33.6827500	-1.9557040	1.4057380
10	ARM30T16	2020-01-01 00:15:00	305.2623000	295.7905000	296.4310000	32.8098800	32.9257200	33.3067600	17.2358500	1.9861990
11	ARM30T19	2020-01-01 00:15:00	None	None	None	None	None	None	None	0.6767570

Figura 35. Datos iniciales

ARM30	
VA_NULOS_INICIALES	33
VA_NULOS_FINALES	0
VB_NULOS_INICIALES	33
VB_NULOS_FINALES	0
VC_NULOS_INICIALES	33
VC_NULOS_FINALES	0
IA_NULOS_INICIALES	33
IA_NULOS_FINALES	33
IB_NULOS_INICIALES	33
IB_NULOS_FINALES	33
IC_NULOS_INICIALES	33
IC_NULOS_FINALES	33
P_NULOS_INICIALES	33
P_NULOS_FINALES	33
Q_NULOS_INICIALES	0
Q_NULOS_FINALES	0

Figura 36. Reporte

Este es un caso donde apenas de tener corrientes perdidas no fue posible, imputarlas puesto que por medición hay mas de un circuito perdido, entonces tocarían hacer suposiciones que terminarían siendo peor que utilizar un modelo de ML o DL.

Propuestas de modelos de machine learning o Deep learning

Como se puede ver, en el diagrama anterior, las dos primeras alternativas para imputar datos es el algoritmo de análisis nodal y cálculo de potencias, por medio del cual se pueden imputar algunas corrientes, algunos voltajes y algunas

potencias, además de que se puede corregir la calidad de cada una de estas variables; tal y como se describe en el algoritmo anterior. Posteriormente, se procederá a hacer un análisis del sistema PRIME para determinar la fiabilidad de imputar datos en los sistemas SCADA a partir del PRIME. Luego, los datos que no se puedan imputar con los anteriores métodos, se tratarán de imputar con modelos clásicos de Machine Learning y modelos de Deep Learning, **teniendo en cuenta que estos modelos deben ser multivariados; ya que entre las variables de interés a imputar existe una alta dependencia.**

Los siguientes son los modelos clásicos de ML más utilizados para modelar series de tiempo multivariadas; por tanto, serán los primeros modelos que se entrenarán y evaluarán:

- Interpolación Lineal o Polinómica.
- K-Nearest Neighbors (KNN).
- Random Forest.
- Máquinas de Vectores de Soporte (SVM).

Posteriormente, se entrenarán y evaluarán las siguientes arquitecturas de Deep Learning, las cuales son muy utilizadas para modelar secuencias:

- Redes Neuronales Recurrentes (RNN). **[POCA MEMORIA]**
- LSTM (Long Short-Term Memory). **[POCA MEMORIA]**
- Arquitecturas Transformer. **[DATOS EN PARALELO]**

Una vez entrenados y evaluados los modelos anteriormente mencionados, se procederá a analizar cuál es el modelo que ofrece un mejor equilibrio entre rendimiento y costo computacional.

También es importante destacar que en el pipeline anterior podría entrar un posible análisis de flujo de potencias.