Datasets and Model Descriptions

**Forest Fires**

Description -

Variables -

Model No. 0 - lm(area ~ (X + Y + month + day + FFMC + DMC + DC + ISI + temp + RH + wind + rain), data = data)

Model No. 1 - lm(area ~ (X + Y + month + day + FFMC + DMC + DC + ISI + temp + RH + wind + rain + I(FFMC^2) + I(DMC^2) + I(DC^2) + I(ISI^2) + I(temp^2) + I(RH^2) + I(wind^2) + I(rain^2)), data = data)

Model No. 2 - lm(area ~ (X + Y + month + day + FFMC + DMC + DC + ISI + temp + RH + wind + rain)^2, data = data)

Model No. 3 - lm(area ~ (X + Y + month + day + FFMC + DMC + DC + ISI + temp + RH + wind + rain + I(FFMC^2) + I(DMC^2) + I(DC^2) + I(ISI^2) + I(temp^2) + I(RH^2) + I(wind^2) + I(rain^2))^2, data = data)

**Wine**

Description - The data is related to white variants of the Portuguese "Vinho Verde" wine. The inputs include objective tests (e.g. PH values) and the output is based on sensory data (median of at least 3 evaluations made by wine experts). Each expert graded the wine quality between 0 (very bad) and 10 (very excellent).

Variables -

Model No. 0 - lm(qual ~ (vol.acid + cit.acid + res.sugar + chlor + free.sulfur + total.sulfur + dens + PH + sulph + alcohol + fixed.acid), data = data)

Model No. 1 – lm(qual ~ (vol.acid + cit.acid + res.sugar + chlor + free.sulfur + total.sulfur + dens + PH + sulph + alcohol + fixed.acid + I(vol.acid^2) + I(cit.acid^2) + I(res.sugar^2) + I(chlor^2) + I(free.sulfur^2) + I(total.sulfur^2) + I(dens^2) + I(PH^2) + I(sulph^2) + I(alcohol^2) + I(fixed.acid^2)), data = data)

Model No. 2 - lm(qual ~ (vol.acid + cit.acid + res.sugar + chlor + free.sulfur + total.sulfur + dens + PH + sulph + alcohol + fixed.acid)^2, data = data)

Model No. 3 - lm(qual ~ (vol.acid + cit.acid + res.sugar + chlor + free.sulfur + total.sulfur + dens + PH + sulph + alcohol + fixed.acid)^3, data = data)

Model No. 4 - lm(qual ~ (vol.acid + cit.acid + res.sugar + chlor + free.sulfur + total.sulfur + dens + PH + sulph + alcohol + fixed.acid + I(vol.acid^2) + I(cit.acid^2) + I(res.sugar^2) + I(chlor^2) + I(free.sulfur^2) + I(total.sulfur^2) + I(dens^2) + I(PH^2) + I(sulph^2) + I(alcohol^2) + I(fixed.acid^2))^2, data = data)

**Student Grades**

Description -

Variables -

Model No. 0 - lm(G1 ~ (sex + age + address + Pstatus + Medu + Fedu + traveltime + studytime + failures + schoolsup + absences),data = data)

Model No. 1 – lm(G1 ~ (sex + age + address + Pstatus + Medu + Fedu + traveltime + studytime + failures + schoolsup + absences + I(age^2) + I(Medu^2) + I(Fedu^2) + I(traveltime^2) + I(studytime^2) + I(failures^2) + I(absences^2)),data = data)

Model No. 2 - lm(G1 ~ (sex + age + address + Pstatus + Medu + Fedu + traveltime + studytime + failures + schoolsup + absences)^2,data = data)

Model No. 3 - lm(G1 ~ (sex + age + address + Pstatus + Medu + Fedu + traveltime + studytime + failures + schoolsup + absences + I(age^2) + I(Medu^2) + I(Fedu^2) + I(traveltime^2) + I(studytime^2) + I(failures^2) + I(absences^2))^2,data = data)

Model No. 4 - lm(G1 ~ (sex + age + address + Pstatus + Medu + Fedu + traveltime + studytime + failures + schoolsup + absences)^3,data = data)

**Parkinson**

Description - This dataset is composed of a range of biomedical voice measurements from 42 people with early-stage Parkinson's disease recruited to a six-month trial of a telemonitoring device for remote symptom progression monitoring. The recordings were automatically captured in the patient's homes.

Variables -

Model No. 0 - lm(motor.updrs ~ (test.time + jitter.percent + jitter.abs + jitter.rap + shim + shim.db + shim.apq3 + shim.dda + nhr + hnr + rdpe + dfa), data = data)

Model No. 1 – lm(motor.updrs ~ (age + sex + test.time + jitter.percent + jitter.abs + jitter.rap + jitter.ppq5 + jitter.ddp + shim + shim.db + shim.apq3 + shim.dda + nhr + hnr + rdpe + dfa), data = data)

Model No. 2 - lm(motor.updrs ~ (test.time + jitter.percent + jitter.abs + jitter.rap + shim + shim.db + shim.apq3 + shim.dda + nhr + hnr + rdpe + dfa+ I(test.time^2) + I(jitter.percent^2) + I(jitter.abs^2) + I(jitter.rap^2) + I(shim^2) + I(shim.db^2) + I(shim.apq3^2) + I(shim.dda^2) + I(nhr^2) + I(hnr^2) + I(rdpe^2) + I(dfa^2)), data = data)

Model No. 3 - lm(motor.updrs ~ (age + sex + test.time + jitter.percent + jitter.abs + jitter.rap + jitter.ppq5 + jitter.ddp + shim + shim.db + shim.apq3 + shim.dda + nhr + hnr + rdpe + dfa + I(age^2)+ I(test.time^2) + I(jitter.percent^2) + I(jitter.abs^2) + I(jitter.rap^2)+ I(jitter.ppq5^2)+ I(jitter.ddp^2) + I(shim^2) + I(shim.db^2) + I(shim.apq3^2) + I(shim.dda^2) + I(nhr^2) + I(hnr^2) + I(rdpe^2) + I(dfa^2)), data = data)

Model No. 4 – lm(motor.updrs ~ (test.time + jitter.percent + jitter.abs + jitter.rap + shim + shim.db + shim.apq3 + shim.dda + nhr + hnr + rdpe + dfa)^2, data = data)

Model No. 5 - lm(motor.updrs ~ (age + sex + test.time + jitter.percent + jitter.abs + jitter.rap + jitter.ppq5 + jitter.ddp + shim + shim.db + shim.apq3 + shim.dda + nhr + hnr + rdpe + dfa)^2, data = data)

**Titanic(glm)**

Description -

Variables -

Model No. 0 - glm(survived ~ (sex + age + pclass + sibsp + parch), data=data, family = "binomial")

Model No. 1 – glm(survived ~ (sex + age + pclass + sibsp + parch + I(age^2)), data=data, family = "binomial")

Model No. 2 - glm(survived ~ (sex + age + pclass + sibsp + parch)^2, data=data, family = "binomial")

Model No. 3 - glm(survived ~ (sex + age + pclass + sibsp + parch + I(age^2))^2, data=data, family = "binomial")

Model No. 4 – glm(survived ~ (sex + age + pclass + sibsp + parch + fare)^2, data=data, family = "binomial")

Model No. 5 - glm(survived ~ (sex + age + pclass + sibsp + parch + fare + I(age^2) + I(fare^2))^2, data=data, family = "binomial")

Model No. 6 - glm(survived ~ (sex + age + pclass + sibsp + parch)^3, data=data, family = "binomial")

Model No. 7 - glm(survived ~ (sex + age + pclass + sibsp + parch + fare)^3, data=data, family = "binomial")

**Power Plant**

Description -

Variables -

Model No. 0 - lm(energy.out ~ (temp + vacuum + ambient.pres + relative.humid), data = data)

Model No. 1 – lm(energy.out ~ (temp + vacuum + ambient.pres + relative.humid + I(temp^2) + I(vacuum^2) + I(ambient.pres^2) + I(relative.humid^2)), data = data)

Model No. 2 - lm(energy.out ~ (temp + vacuum + ambient.pres + relative.humid)^2, data = data)

Model No. 3 - lm(energy.out ~ (temp + vacuum + ambient.pres + relative.humid)^3, data = data)

Model No. 4 – lm(energy.out ~ (temp + vacuum + ambient.pres + relative.humid)^4, data = data)

Model No. 5 - lm(energy.out ~ (temp + vacuum + ambient.pres + relative.humid + I(temp^2) + I(vacuum^2) + I(ambient.pres^2) + I(relative.humid^2))^2, data = data)

Model No. 6 - lm(energy.out ~ (temp + vacuum + ambient.pres + relative.humid + I(temp^2) + I(vacuum^2) + I(ambient.pres^2) + I(relative.humid^2))^3, data = data)

Model No. 7 - lm(energy.out ~ (temp + vacuum + ambient.pres + relative.humid + I(temp^2) + I(vacuum^2) + I(ambient.pres^2) + I(relative.humid^2))^4, data = data)

**News Popularity**

Description -

Variables -

Model No. 0 - lm(shares ~ (n\_tokens\_title + n\_tokens\_content + n\_unique\_tokens + n\_non\_stop\_words + n\_non\_stop\_unique\_tokens + num\_hrefs), data = data)

Model No. 1 – lm(shares ~ (n\_tokens\_title + n\_tokens\_content + n\_unique\_tokens + n\_non\_stop\_words + n\_non\_stop\_unique\_tokens + num\_hrefs + num\_self\_hrefs + num\_imgs + num\_videos + average\_token\_length + num\_keywords), data = data)

Model No. 2 - lm(shares ~ (n\_tokens\_title + n\_tokens\_content + n\_unique\_tokens + n\_non\_stop\_words + n\_non\_stop\_unique\_tokens + num\_hrefs + I(n\_tokens\_title^2) + I(n\_tokens\_content^2) + I(n\_unique\_tokens^2) + I(n\_non\_stop\_words^2) + I(n\_non\_stop\_unique\_tokens^2) + I(num\_hrefs^2)), data = data)

Model No. 3 - lm(shares ~ (n\_tokens\_title + n\_tokens\_content + n\_unique\_tokens + n\_non\_stop\_words + n\_non\_stop\_unique\_tokens + num\_hrefs + num\_self\_hrefs + num\_imgs + num\_videos + average\_token\_length + num\_keywords + I(n\_tokens\_title^2) + I(n\_tokens\_content^2) + I(n\_unique\_tokens^2) + I(n\_non\_stop\_words^2) + I(n\_non\_stop\_unique\_tokens^2) + I(num\_hrefs^2) + I(num\_self\_hrefs^2) + I(num\_imgs^2) + I(num\_videos^2) + I(average\_token\_length^2) + I(num\_keywords^2)), data = data)

Model No. 4 – lm(shares ~ (n\_tokens\_title + n\_tokens\_content + n\_unique\_tokens + n\_non\_stop\_words + n\_non\_stop\_unique\_tokens + num\_hrefs)^2, data = data)

Model No. 5 - lm(shares ~ (n\_tokens\_title + n\_tokens\_content + n\_unique\_tokens + n\_non\_stop\_words + n\_non\_stop\_unique\_tokens + num\_hrefs)^3, data = data)

Model No. 6 - lm(shares ~ (n\_tokens\_title + n\_tokens\_content + n\_unique\_tokens + n\_non\_stop\_words + n\_non\_stop\_unique\_tokens + num\_hrefs + I(n\_tokens\_title^2) + I(n\_tokens\_content^2) + I(n\_unique\_tokens^2) + I(n\_non\_stop\_words^2) + I(n\_non\_stop\_unique\_tokens^2) + I(num\_hrefs^2))^2, data = data)

Model No. 7 – lm(shares ~ (n\_tokens\_title + n\_tokens\_content + n\_unique\_tokens + n\_non\_stop\_words + n\_non\_stop\_unique\_tokens + num\_hrefs + num\_self\_hrefs + num\_imgs + num\_videos + average\_token\_length + num\_keywords)^2, data = data)

Model No. 8 - lm(shares ~ (n\_tokens\_title + n\_tokens\_content + n\_unique\_tokens + n\_non\_stop\_words + n\_non\_stop\_unique\_tokens + num\_hrefs + I(n\_tokens\_title^2) + I(n\_tokens\_content^2) + I(n\_unique\_tokens^2) + I(n\_non\_stop\_words^2) + I(n\_non\_stop\_unique\_tokens^2) + I(num\_hrefs^2))^3, data = data)

Model No. 9 - lm(shares ~ (n\_tokens\_title + n\_tokens\_content + n\_unique\_tokens + n\_non\_stop\_words + n\_non\_stop\_unique\_tokens + num\_hrefs + num\_self\_hrefs + num\_imgs + num\_videos + average\_token\_length + num\_keywords)^3, data = data)

Model No. 10 - lm(shares ~ (n\_tokens\_title + n\_tokens\_content + n\_unique\_tokens + n\_non\_stop\_words + n\_non\_stop\_unique\_tokens + num\_hrefs + num\_self\_hrefs + num\_imgs + num\_videos + average\_token\_length + num\_keywords + I(n\_tokens\_title^2) + I(n\_tokens\_content^2) + I(n\_unique\_tokens^2) + I(n\_non\_stop\_words^2) + I(n\_non\_stop\_unique\_tokens^2) + I(num\_hrefs^2) + I(num\_self\_hrefs^2) + I(num\_imgs^2) + I(num\_videos^2) + I(average\_token\_length^2) + I(num\_keywords^2))^2, data = data)

**Concrete**

Description - The actual concrete compressive strength (MPa) for a given mixture under a specific age (days) was determined from laboratory.

Variables -

Model No. 0 - lm(cement ~ (blast + fly.ash + water + superplas), data = data)

Model No. 1 – lm(cement ~ (blast + fly.ash + water + superplas + I(blast^2) + I(fly.ash^2) + I(water^2) + I(superplas^2)), data = data)

Model No. 2 - lm(cement ~ (blast + fly.ash + water + superplas)^2, data = data)

Model No. 3 - lm(cement ~ (blast + fly.ash + water + superplas + coarse.agg + fine.agg + age + compress.str), data = data)

Model No. 4 – lm(cement ~ (blast + fly.ash + water + superplas)^3, data = data)

Model No. 5 - lm(cement ~ (blast + fly.ash + water + superplas + coarse.agg + fine.agg + age + compress.str + I(blast^2) + I(fly.ash^2) + I(water^2) + I(superplas^2) + I(coarse.agg^2) + I(fine.agg^2) + I(age^2) + I(compress.str^2)), data = data)

Model No. 6 - lm(cement ~ (blast + fly.ash + water + superplas + coarse.agg + fine.agg + age + compress.str)^2, data = data)

Model No. 7 – lm(cement ~ (blast + fly.ash + water + superplas + I(blast^2) + I(fly.ash^2) + I(water^2) + I(superplas^2))^2, data = data)

Model No. 8 - lm(cement ~ (blast + fly.ash + water + superplas + coarse.agg + fine.agg + age + compress.str)^3, data = data)

Model No. 9 - lm(cement ~ (blast + fly.ash + water + superplas + I(blast^2) + I(fly.ash^2) + I(water^2) + I(superplas^2))^3, data = data)

Model No. 10 - lm(cement ~ (blast + fly.ash + water + superplas + coarse.agg + fine.agg + age + compress.str + I(blast^2) + I(fly.ash^2) + I(water^2) + I(superplas^2) + I(coarse.agg^2) + I(fine.agg^2) + I(age^2) + I(compress.str^2))^2, data = data)

Model No. 11 - lm(cement ~ (blast + fly.ash + water + superplas + coarse.agg + fine.agg + age + compress.str)^4, data = data)

**Facebook**

Description - The data is related to posts published during the year of 2014 on the Facebook's page of a renowned cosmetics brand.

Variables -

Model No. 0 - lm(Y ~ (X1C + X2C + X3C + X4C + X5C + X6C), data = data)

Model No. 1 – lm(Y ~ (V01 + V02 + V03 + V04 + V05 + V06 + V07 + V08 + V09 + V10 + V11 + V12), data = data)

Model No. 2 - lm(Y ~ (X1C + X2C + X3C + X4C + V01 + V02 + V03 + V04 + V05 + V06 + V07), data = data)

Model No. 3 - lm(Y ~ (X1C + X2C + X3C + V01 + V02 + V03 + V04 + V05 + V06 + V07 + V08 + V09), data = data)

Model No. 4 – lm(Y ~ (X1C + X2C + X3C + X4C + X5C + X6C + V01 + V02 + V03 + V04 + V05 + V06 + V07 + V08 + V09 + V10 + V11 + V12), data = data)

Model No. 5 - lm(Y ~ (V01 + V02 + V03 + V04 + V05 + V06 + V07 + V08 + V09 + V10 + V11 + V12 + I(V01^2) + I(V02^2) + I(V03^2) + I(V04^2) + I(V05^2) + I(V06^2) + I(V07^2) + I(V08^2) + I(V09^2) + I(V10^2) + I(V11^2) + I(V12^2)), data = data)

Model No. 6 - lm(Y ~ (X1C + X2C + X3C + X4C + V01 + V02 + V03 + V04 + V05 + V06 + V07 + I(V01^2) + I(V02^2) + I(V03^2) + I(V04^2) + I(V05^2) + I(V06^2) + I(V07^2)), data = data)

Model No. 7 – lm(Y ~ (X1C + X2C + X3C + V01 + V02 + V03 + V04 + V05 + V06 + V07 + V08 + V09 + I(V01^2) + I(V02^2) + I(V03^2) + I(V04^2) + I(V05^2) + I(V06^2) + I(V07^2) + I(V08^2) + I(V09^2)), data = data)

Model No. 8 - lm(Y ~ (X1C + X2C + X3C + X4C + V01 + V02 + V03 + V04 + V05 + V06 + V07 + V08 + V09 + V10 + V11 + V12 + I(V01^2) + I(V02^2) + I(V03^2) + I(V04^2) + I(V05^2) + I(V06^2) + I(V07^2) + I(V08^2) + I(V09^2) + I(V10^2) + I(V11^2) + I(V12^2)), data = data)

Model No. 9 - lm(Y ~ (X1C + X2C + X3C + X4C + X5C + X6C + V01 + V02 + V03 + V04 + V05 + V06 + V07 + V08 + V09 + V10 + V11 + V12 + I(V01^2) + I(V02^2) + I(V03^2) + I(V04^2) + I(V05^2) + I(V06^2) + I(V07^2) + I(V08^2) + I(V09^2) + I(V10^2) + I(V11^2) + I(V12^2)), data = data)

Model No. 10 - lm(Y ~ (V01 + V02 + V03 + V04 + V05 + V06 + V07 + V08 + V09 + V10 + V11 + V12)^2, data = data)

Model No. 11 – lm(Y ~ (X1C + X2C + X3C + X4C + X5C + X6C)^2, data = data)

Model No. 12 - lm(Y ~ (X1C + X2C + X3C + V01 + V02 + V03 + V04 + V05 + V06 + V07 + V08 + V09)^2, data = data)

Model No. 13 - lm(Y ~ (X1C + X2C + X3C + X4C + V01 + V02 + V03 + V04 + V05 + V06 + V07)^2, data = data)

Model No. 14 - lm(Y ~ (X1C + X2C + X3C + X4C + V01 + V02 + V03 + V04 + V05 + V06 + V07 + V08 + V09 + V10 + V11 + V12)^2, data = data)

Model No. 15 - lm(Y ~ (V01 + V02 + V03 + V04 + V05 + V06 + V07 + V08 + V09 + V10 + V11 + V12)^3, data = data)

Model No. 16 - lm(Y ~ (V01 + V02 + V03 + V04 + V05 + V06 + V07 + V08 + V09 + V10 + V11 + V12 + I(V01^2) + I(V02^2) + I(V03^2) + I(V04^2) + I(V05^2) + I(V06^2) + I(V07^2) + I(V08^2) + I(V09^2) + I(V10^2) + I(V11^2) + I(V12^2))^2, data = data)

Model No. 17 - lm(Y ~ (X1C + X2C + X3C + V01 + V02 + V03 + V04 + V05 + V06 + V07 + V08 + V09 + I(V01^2) + I(V02^2) + I(V03^2) + I(V04^2) + I(V05^2) + I(V06^2) + I(V07^2) + I(V08^2) + I(V09^2))^2, data = data)

Model No. 18 - lm(Y ~ (X1C + X2C + X3C + X4C + X5C + X6C)^3, data = data)